

NBER WORKING PAPER SERIES

NOISE IN EXPECTATIONS:
EVIDENCE FROM ANALYST FORECASTS

Tim de Silva
David Thesmar

Working Paper 28963
<http://www.nber.org/papers/w28963>

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
June 2021

We thank Alberto Abadie, Sam Anderson, Anne Beyer, Harin de Silva, Eric Ghysels, Allen Hu, Eben Lazarus, Charles Lee, Eric So, Frank Schilbach, Dmitry Taubinsky, and seminar and conference participants. An earlier version of this paper were circulated with the title “The Term Structure of Subjective Expectations: Evidence and Theory from Analysts Forecasts”. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2021 by Tim de Silva and David Thesmar. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

Noise in Expectations: Evidence from Analyst Forecasts
Tim de Silva and David Thesmar
NBER Working Paper No. 28963
June 2021
JEL No. D84,D9,D91

ABSTRACT

This paper quantifies the amount of noise and bias in analysts' forecast of corporate earnings at various horizons. We first show analyst forecasts outperform statistical forecasts at short-horizons, but underperform at longer horizons. We next decompose the relative accuracy of these forecasts into three components: (i) noise, (ii) bias and (iii) analysts' information advantage over statistical forecasts. We find the information advantage is constant across forecasting horizons, while both noise and bias are increase linearly. We then show most existing models lack a mechanism to account for these facts. To generate such a mechanism, we consider a parsimonious variant of the model of Patton and Timmermann (2010) with a noisy cognitive default and show it quantitatively fits the data. The intuition underlying this model is that forecasters rely on their biased and noisy defaults more at longer horizons, as rational forecasts are less accurate. This model also quantitatively matches two non-targeted empirical relationships: (i) analyst disagreement increases with horizon and (ii) noise is an increasing function of volatility.

Tim de Silva
MIT Sloan School of Management
tdesilva@mit.edu

David Thesmar
MIT Sloan School of Management
100 Main Street, E62-632
Cambridge, MA 02142
and NBER
thesmar@mit.edu

It is well-known that human judgment is noisy, meaning two individuals with the same information may make meaningfully different predictions or decisions. In a recent book, [Kahneman, Sibony, and Sunstein \(2021\)](#) discuss many areas in which noise is prevalent, including medicine, finance, hiring, and the justice system. In this paper, we study noise in equity analysts' corporate earnings forecasts. We begin by quantifying it, distinctly from forecasting bias (the focus of most existing literature), and analyzing its term structure. We next explore ways to adjust for forecasting noise, and finally propose a model of its origin that is quantitatively consistent with the evidence we document.

Our approach to quantifying noise relies on a simple decomposition. Denote $F_{it}y_{t+1}$ as the forecast of y_{t+1} formed by forecaster i at date t and I_{it} as her information set. Additionally, denote X_t as the information set of an econometrician. We can rewrite the forecaster's forecast as:

$$\underbrace{F_{it}y_{t+1}}_{\text{subjective forecast}} = \underbrace{E(y_{t+1}|X_t)}_{\text{econometric forecast}} + \underbrace{E(y_{t+1}|I_{it}) - E(y_{t+1}|X_t)}_{\text{forecaster's private information}} + \underbrace{B(I_{it})}_{\text{bias}} + \underbrace{\eta_{it}}_{\text{noise}},$$

where by definition $B(I_{it}) = E(F_{it}y_{t+1} - y_{t+1}|I_{it})$ and thus $\eta_{it} = F_{it}y_{t+1} - E(F_{it}y_{t+1}|I_{it})$. The bias term in this equation is the predictable deviation from rational expectations, conditional on available information. Noise comes on top of the bias, private information, and the rational component – it is outside of the information set by definition. Now consider the econometrician's forecast using X_t , and further assume that $X_t \subset I_{it}$. Under mild assumptions, the econometric MSE writes:

$$\underbrace{MSE}_{\text{subjective MSE}} = \underbrace{MSE^m}_{\text{econometric MSE}} - \underbrace{E(E(y_{t+1}|I_{it}) - E(y_{t+1}|X_t))^2}_{\text{forecaster's information advantage}} + \underbrace{EB(I_{it})^2}_{\text{bias}} + \underbrace{var(\eta_{it})}_{\text{noise}}, \quad (1)$$

where $MSE^m = E(y_{t+1} - E(y_{t+1}|X_t))^2$ is the econometric MSE (i.e. the rational MSE given X_t). According to (1), the difference in MSE between subjective forecasters and the econometrician can be decomposed into three components: noise, bias, and the forecasters' informational advantage (as in [Satopää, Salikhov, Tetlock, and Mellers 2020](#)).

This paper uses the above decomposition, along with some additional assumptions, to quantify the noise and bias in equity analysts forecasts of corporate earnings. Analyst forecast data are an interesting in several respects: equity analysts are skilled and incentivized forecasters, and forecast data are available for a large panel of firms at various horizons. Beyond analyst forecasts, our approach is simple and applicable to any dataset on subjective forecasts. A key difference from existing literature is our approach does not make strong

restrictions on the data-generating process.

First, in order to perform the decomposition described above to estimate expectation noise, we need to compare subjective and econometric expectations. To form econometric forecasts, we explore a variety of well-known supervised machine learning estimators (e.g. Lasso, Ridge, and tree-based methods). We find these estimators clearly dominate simpler OLS approaches, consistent with existing evidence (e.g. Ball and Ghysels 2018; van Binsbergen, Han, and Lopez-Lira 2020; Cao, Jiang, Wang, and Yang 2021). At the one-year horizon, econometric forecasts are less precise than subjective ones, but this dramatically reverses at longer horizons.

We then quantify the three components that determine the difference in accuracy between analyst and econometric forecasts: (i) information advantage; (ii) bias; (iii) noise. In contrast to most of the extant literature, we place no restriction on the data generating process. Our approach does rely on one key assumption on the belief formation process: forecasters are *unbiased* with respect to their private information. Under this assumption, we show that estimating the three components of the decomposition in (1) is simple.

Applying our estimation strategy, we find that noise in subjective expectations is large. Quantitatively, we estimate the amount of noise in analyst expectations is about 60% of the magnitude of the bias (i.e. the predictable component of analyst forecast errors), which is especially large given we work with consensus forecasts. At the one-year horizon, we find forecaster bias and noise are small compared to the analyst information advantage – this is why short-horizon subjective expectations dominate econometric forecasts. At longer forecasting horizons (two and three-years), information advantage does not decay, but forecaster bias and noise increase dramatically. Quantitatively, bias is more than twice as large as the information advantage at the three-year horizon and noise is 1.5 times bigger. In sum, forecaster bias and noise both sharply increase with horizon, while the term structure of information advantage is flat.

Given noise and bias deteriorate the quality of subjective forecasts (as the (1) makes clear), our findings suggest two ways of improving the predictive power of forecasts. The first approach consists of an optimal combination of statistical and subjective forecasts. We implement such a combination using the same supervised learning approaches as above and find penalized-linear methods are slightly dominated by tree-based methods. At short horizons, the improvement from using this optimal combined forecast relative to subjective forecasts is small – a meak 10% reduction of MSE – which is expected given these forecasts

have little bias and noise. However, since noise and bias are much larger at longer horizons, there is a much larger improvement from optimal combined forecast – a reduction in MSE of about 33%. These findings are reminiscent of the recommendation from [Kahneman et al. \(2021\)](#) to use mechanical rules when human judgment is noisy. In this setting, the combined forecast implements a mechanical rule in the form of shrinkage: subjective expectations are shrunk towards the statistical rule, which reduces noise at the expense of losing information.

The second approach to tame noise and improve precision is increasing the number of forecasters in the consensus – the classic “wisdom of crowds” argument. In the data, we find when there is a larger number of analysts, the consensus exhibits a smaller amount of noise, but noise does not vanish to zero as the number of analysts grows very large. The success of this aggregation suggests noise is indeed imperfectly correlated across forecasters, but there might be some non-zero correlation.

Next, we examine which models of belief formation can *jointly* match the upward sloping term structures of noise and bias we estimate. We first revisit a large class of models, including models of noisy information ([Woodford 2003](#)), bounded rationality ([Sims 2003](#)), diagnostic expectations ([Bordalo, Gennaioli, Ma, and Shleifer 2020](#)), over-confidence ([Daniel, Subrahmanyam, and Hirshleifer 1998](#)), and over-extrapolation ([Greenwood and Shleifer 2014](#); [Angeletos, Huo, and Sastry 2020](#)). In their simple form, all these models predict downward sloping term structures for both bias and noise. This is because these models rely on the law of iterated expectations to determine the term structure of forecasts: since forecasters know the true data-generating process, they recognize the best forecast for long horizons is the unconditional mean, which is the rational expectation. These models essentially assume reversion in forecasts to the true mean as the forecasting horizon increases.

Motivated by the failure of these models, we deviate by exploring a variant of the model in [Patton and Timmermann \(2010\)](#) that has two key components. First, forecasters exhibit a form of bounded rationality in the spirit of [Gabaix \(2014\)](#). Specifically, forecasts are a weighted sum of a cognitive default and the true conditional expectation, with less weight on the former as the latter becomes more accurate. This dependence is assumed, but not micro-founded following [Patton and Timmermann \(2010\)](#). The second key ingredient is the cognitive default of forecasters is anchored on the current observation (e.g. [Kahneman and Tversky 1972](#)), but also contains cognitive noise. This model has three key horizon-invariant parameters: one for default bias, one for default noise, and one that determines the relative weight on the anchor.

We then formally estimate this model and find it quantitatively matches the term structures of bias and noise we document. The estimated noise in cognitive defaults is twice as large as the variation in the true data generating process, which allows us to match the large average level of noise in the data. We also find evidence of a tendency of analysts to overweight the current observation, which helps match the average level of bias. Our ability to match the upward *slope* of the bias and noise term structures is driven by a form of bounded rationality: forecasters rely more on their cognitive defaults at longer horizons because the true conditional expectation is less accurate in absolute terms. Quantitatively, the level of bounded rationality we estimate is similar to what the attention function calibrated in Gabaix (2019) would predict in our setting.

Finally, we explore two additional predictions of the model. The first prediction is that forecaster disagreement should increase with horizon – the original variable of interest in Patton and Timmermann (2010) – as all disagreement in the model arises from noisy cognitive defaults. Although the model parameters were matched on the term structure of noise and bias, we find the predicted term structure of disagreement matches the data well. The second prediction of the model is that noise is increasing with earnings volatility. The intuition is forecasters choose to rely more on their noisy defaults at all horizons when earnings are harder to forecast. When sorting firms based on volatility, we find the model quantitatively matches the empirical relation between noise and residual volatility.

Related literature. Subjective forecast noise is discussed in the large literature on noisy information (e.g. Woodford 2003; Azeredo da Silveira, Sung, and Woodford 2020) and in cognitive psychology (e.g. Khaw, Li, and Woodford 2019; Woodford 2020; Kahneman et al. 2021). Our contribution to this literature is twofold: (i) our evidence on the size and term structure of noise (in contrast to bias) using analyst forecast data and (ii) our methodology that places no restrictions on the data-generating process. Our methodology is similar in spirit to Satopää et al. (2020), who perform a similar bias-information-noise (“BIN”) decomposition and find a consistent property of good subjective forecasters is noise reduction, consistent with our proposals for noise reduction. Our approach is also related to Bianchi, Ludvigson, and Ma (2020) and Nagel (2021), who discuss how supervised learning is useful for studying subjective expectations data.

Our work is also related to the extant empirical literature on expectations formation. This literature generally focuses on estimating forecaster bias (e.g. Manski 2017) and forecaster information *dis*advantage (e.g. Coibion and Gorodnichenko 2015). In contrast, we measure two additional components: subjective forecasters’ information advantage and noise. We

further document the term structure of these components and explore a modeling assumption – reliance on noisy default – that allows us to fit the data.

Our finding of an upward sloping term structure of noise is, to our knowledge, novel. Patton and Timmermann (2010) document that disagreement in macro forecasts increases with horizon, which is consistent with this observation. The upward sloping term-structure of bias is consistent with emerging evidence in asset prices and expectations data (Giglio and Kelly 2018; Bordalo, Gennaioli, La Porta, and Shleifer 2019; D’Arienzo 2020; Angeletos et al. 2020; Afrouzi, Kwon, Landier, Ma, and Thesmar 2021). More precisely, this literature documents more overreaction at long horizons, which our approach cannot show due to our lack of restrictions on the DGP. Closely related evidence is presented in Dessaint, Foucault, and Frésard (2020), who show that long-term forecasts are less predictive of future earnings realization. This is consistent with long-term forecasts being either more biased or noisier (or both).

Because we estimate statistical forecasts, our paper also engages with the recent literature applying supervised machine learning in economics and finance (see Mullainathan and Spiess 2017, for a review). To perform our decomposition, we study the predictability of corporate earnings at various horizons using firm-level observables (from accounting and price data) and standard ML estimators (penalized methods and tree-based methods). So (2013) proposes a parsimonious model to forecast EPS based on OLS. A slew of recent papers have applied ML techniques to this problem, mostly at shorter horizons (see Ball and Ghysels 2018; van Binsbergen et al. 2020; Hansen and Thimsen 2020; Cao and You 2020). Like these papers, and like papers implementing a similar exercise on equity returns directly (e.g. Gu, Kelly, and Xiu 2018; Kozak, Nagel, and Santosh 2020; Bryzgalova, Huang, and Julliard 2020), we find that there are gains to using supervised ML techniques over non-regularized estimators (e.g. OLS). Another outcome of our analysis is that tree-based forecasts marginally dominate penalized methods – this is also consistent with the literature on EPS forecasting.

Finally, our paper is related to the extensive literature on analyst forecasts (see Kothari, So, and Verdi 2016, for a review) in finance and accounting. Our findings that analyst forecasts are more accurate at a horizon of less than a year is broadly consistent with this literature (e.g. Brown and Rozeff 1978; Bradshaw, Drake, Myers, and Myers 2012), and our estimate of a large analyst information advantage corroborates the survey evidence in Brown, Call, Clement, and Sharp (2015). Our proposed model features a form of bounded rationality, consistent with evidence of attention constraints shaping analyst forecast behavior by affecting effort allocation (Harford, Jiang, Wang, and Xie 2019) and inducing social learning

(Kumar, Rantala, and Xu 2021).

Outline. Section 1 describes data and forecast formation. Section 2 compares the accuracy of statistical and subjective forecasts at different horizons. Section 3 introduces and performs our decomposition and subsequently discusses noise reduction. Section 4 explains how existing models fail to match the term structure of bias and noise we document and estimates our proposed model. Section 5 quantitatively tests additional predictions of our model. Section 6 concludes and an [Appendix](#) contains additional results and derivations.

1 Data and Forecast Construction

In this section, we first describe the construction of the data we use. Next, we explain how we build the three forecasts of EPS that form the core of our analysis: (i) the analyst forecast (the “subjective” forecast), (ii) the econometric/statistical forecast (the “machine” forecast), and (iii) the combined forecast (the “machine + analyst” forecast).

1.1 Data Description

The data used in this paper comes from three sources: I/B/E/S, Compustat, and CRSP. We start by collecting the reported fiscal-year-end earnings-per-share (EPS) and their respective announcement dates from the I/B/E/S actuals file for all US firms with announcements between 1990 and 2020. In our analysis, we focus entirely on fiscal-year end earnings reports to avoid issues associated with seasonality in earnings (see [Kothari 2001](#)). We winsorize the current year’s EPS within each year at 5%-95%. Following [Bouchaud, Krüger, Landier, and Thesmar \(2019\)](#), we keep all EPS forecasts for the following three fiscal year ends that are issued within 45 days of the most recent annual earnings announcement (prior fiscal year end). The corresponding realizations of EPS are denoted EPS_{it+h} , where $h \in \{1, 2, 3\}$ years. Ideally we would consider longer forecasting horizons, but the data are insufficiently well-populated. We drop observations where there is more than a 23 month gap before the next announcement.

Next, we merge the remaining observations with Compustat Annual and collect the variables reported in Panel A of [Appendix Table A1](#) at the fiscal year end for firm i in year t . These are the most common Compustat variables, which we chose in order to maximize the

number of variables that had a large number of non-missing observations. We use Compu-stat data instead of ratios identified by prior literature as good predictors (e.g. Hansen and Thimsen 2020) in order to mitigate look ahead bias in our statistical forecasts, which would occur if we used variables identified ex-post by prior literature. We then merge the remaining data with CRSP and collect variables listed in Panel B of Table A1, making adjustments for performance-related delistings (Shumway 1997). Finally, we delete all observations for securities that are not ordinary equity securities (CRSP share codes 10 and 11).

We impose two final sample filters. First, we require that a firm-year (i, t) has a non-missing value of total assets in year $t-1$ and all checked variables in Table A1 are non-missing. Secondly, we delete all firms classified as financial firms by their 1-digit SIC code.¹

1.2 Analyst Forecasts

We denote $F_t EPS_{it+h}$ as the consensus of analyst EPS forecasts of firm i for year $t+h$. More precisely, we compute the median² of analyst forecasts for fiscal year $t+h$, issued within 45 days after the announcement of annual earnings t . We only keep the earliest forecast of each analyst when calculating the median, if an analyst issues multiple forecasts. We focus on this 45 days period to make sure our subjective forecasts are not stale and are taken with similar information sets across analysts³, but also to make the information set contemporaneous with the machine forecast, to which we now turn.

1.3 Machine Forecasts

We denote $F_t^m EPS_{it+h}$ as the machine forecast for firm i as of date t for fiscal year $t+h$. Given a set of public information X_{it} , our goal is to approximate the conditional expectation function, $E(EPS_{it+h}|X_{it})$, as accurately as possible. As is well-known, $E(EPS_{it+h}|X_{it})$ is the solution to the problem of minimizing mean squared error across all possible (measurable)

¹In addition, we delete a very small number observations where the median analyst forecast error normalized by price in our sample of EPS_{it+h} is outside 8 times the interquartile range, which are likely data errors. Aside from this final criterion, none of our sample filters described in this section induce look-ahead bias, which is important to ensure comparability in information sets between subjective and statistical forecasts.

²We choose to use the median analyst forecast to mitigate the impact of a few outliers, but we note the correlation between the mean and median analyst forecast in our sample is 0.9993.

³Our main results are robust to using a 30 day instead of 45 day window for forecasts.

functions of X_{it} :

$$\begin{aligned} E_t(EP S_{it+h}|X_{it}) &= \arg \min_{h(X_{it})} E[(EP S_{it+h} - h(X_{it}))^2] \\ &\equiv f_t(X_{it}) \end{aligned} \tag{2}$$

In practice, solving (2) is infeasible because it requires searching over an infinite dimensional function space. To gain tractability, we leverage supervised machine learning approaches that restrict $h(X_{it})$ to be within a particular class of functions, such as linear functions, and use different forms of regularization developed in supervised machine learning to address the high dimensionality of the set of variables in X_{it} . Before describing the details of these procedures, we first describe how we define our firm-level conditioning variables, which constitute the information set X_{it} .

To form X_{it} , we use all the variables listed in Table A1, which come from Compustat and CRSP.⁴ We scale all of the Compustat variables by total assets except total assets itself, EPS, book equity per share, SIC codes, end-of-year price per share from Compustat. We use these variables for fiscal year t , $t-1$ and $t-2$. Given that we forecast EPS for year $t+h$, this means that the machine needs to wait until the annual earnings announcements in year t before forming its forecast. So the machine forecast can be thought of as being issued at the same time as the analyst forecast described above.

To solve (2) and generate an estimate of $f_t(\cdot)$, we next need to define the training sample. To avoid look-ahead bias, we use rolling windows of 5 years to train the various statistical forecasting models. This period is chosen to maximize the size of the training set subject to computational constraints.⁵ Moreover, by choosing a rolling window, we implicitly allow for low-frequency changes in the data generating process for earnings over time. More precisely, for each date t , we use all firm-year observations (i, s) in years $s \in \{t-4, t-3, \dots, t\}$ to forecast earnings at $t+1, t+2, t+3$. Given forecasting variables X_{is} , our estimation consists in finding the function $f_t(X_{is})$ that has the minimum MSE in explaining in-sample future $EP S_{i,s+h}$, using various regularization techniques described below.

Finally, we describe the supervised learning techniques we use to estimate the forecasting

⁴In Appendix G, we our one-year horizon forecasting results are robust to using an alternative set of variables in X_{it} that are used by van Binsbergen et al. (2020).

⁵To check that our conclusions are not sensitive to the choice of a 5 year window, we ran one of our estimators (Gradient-Boosted Trees) using a growing window, where all past data is used (this is too computationally challenging for our penalized linear estimators, as we include many interactions in those estimations). We found the MSE of the machine forecast reduced only by 2.0%, and the machine + analyst forecast MSE increased by 0.6%.

function $f_t(\cdot)$ by solving (2). We choose to depart from using OLS to form these forecasts for two reasons. First, since our goal is approximation of a conditional expectation functions, we want to impose minimal restrictions on the form of $f_t(X_{it})$. However, if X_{it} was low-dimensional, we could in principle use OLS, which we do explore below as a benchmark using a subset of X_{it} . Secondly, since X_{it} is not low-dimensional, OLS is inconsistent and unstable due to its tendency to overfit. Thus, we turn to supervised learning techniques, which restrict the spaces of possible functions in (2) to be tractable yet flexible, while simultaneously minimizing the risk of over-fitting using various forms of regularization. The following is a brief presentation of our approach; we refer the reader to Appendix A for more discussion on the theoretical properties of these estimators and our implementation.

Simple benchmarks. We consider two simple benchmarks in order to illustrate the gains from ML techniques. The first benchmark we consider is a random walk benchmark, in $F_t^m EPS_{it+h} = EPS_{it}, \forall h \in \{1, 2, 3\}$. Although this will be our worst performing forecast, it is benchmark commonly used in the extensive literature on analyst forecasts (e.g. Bradshaw et al. 2012). Secondly, we consider the procedure developed by So (2013). This methodology uses a restricted number of accounting ratios and is estimated using annual cross-sectional OLS regressions.

Penalized linear estimators. The first set of estimators we explore falls within the class of penalized linear estimators. Given we are interested in capturing non-linearities, we include all two-way interactions (including squares) between variables in X_{it} as features when using penalized linear estimators. This results in approximately 8,000 features in each year. Penalized linear estimators consists in finding the linear combination of these features that minimizes MSE while minimizes over-fitting out of sample.

We explore five penalized linear estimators. The first three are commonly used: Lasso, Ridge, and Elastic Net. These estimators are defined by the solution to the same objective function that OLS solves (minimizing in-sample mean squared error), but with an additional penalty term on the size of the coefficients, where the size of the penalty is chosen via cross-validation on the training set (detailed in Appendix A). Intuitively, cross-validation consists of breaking up the training sample into smaller datasets, fitting models on these smaller datasets, and examining which penalty value generates the best performance on the other parts of the training set. Importantly, cross-validation is done entirely using the training set to avoid introducing any look-ahead bias.

The second two penalized linear estimators we explore are Post-Lasso and Iterative-Lasso.

Post-Lasso consists of running a second-stage OLS regression using only variables with non-zero coefficients in a first-stage Lasso estimation. Post-Lasso has been shown to improve performance, given that Lasso suffers from substantial regularization bias (Chernozhukov, Chetverikov, Demirer, Duflo, Hansen, Newey, and Robins 2016). Iterative-Lasso is calculated by solving the same optimization problem as Lasso, but an alternative iterative method (described in Appendix A) is used to estimate the penalty parameter instead of cross-validation. This iterative method has been shown to generate near-oracle rates of convergence (Belloni, Chernozhukov, and Hansen 2011). Additionally, this iterative method substantially reduces the likelihood of any over-fitting relative to cross-validation, as the penalty parameter is chosen without regard to model performance on the training set.

Tree-based methods. We also consider two tree-based methods: Random Forests (RF) and Gradient-Boosted Trees (GBT). The building block of tree-based estimators are regression trees, which are nonparametric (unlike penalized linear estimators) regression estimators designed to capture arbitrary non-linearities among the variables in X_{it} . Used alone, regression trees have a tendency to over-fit, which led to the development various “ensemble” methods that introduce forms of regularization. RF and GBT are two such ensemble methods based on the same core idea: grow a large number of uncorrelated trees and then average their predictions.

RF is constructed based on the intuition of bootstrapping. On each bootstrapped sample, a regression tree is grown. After doing this multiple times, final predictions from the Random Forest are calculated by averaging predictions across the multiple regression trees. Averaging across many trees, which have different structures due to the randomness in the subset of predictor variables chosen, is the regularization in this method that limits over-fitting and reduces prediction variance. Similar to the penalized linear estimators, the parameters that govern the shape of the regression trees can be chosen using cross-validation on the training set.

GBT starts by fitting a very shallow tree, meaning only a small number of variables are used. This shallow tree is likely has terrible in-sample fit. To improve its fit, a second shallow tree is fit on the residuals calculated from the first tree. Predicted values are then formed by using a weighted average of the predicted values from the two trees. This procedure is repeated many times, after which the predicted value will be a weighted-average of the predicted value from all the shallow trees. The sequential growing of trees on residuals from the previous trees makes the trees less correlated, which is why averaging over trees limits over-fitting. As with the other methods, the parameters governing the size of these shallow

trees and the weights in the weighted average are be chosen using cross-validation on the training set.

1.4 Combined Machine & Analyst Forecasts

Finally, we denote $F_t^{m+a}EPS_{it+h}$ as the forecast that combines the conditioning variables in X_{it} as well as analyst forecasts F_tEPS_{it+h} . Both types of information are available to the econometrician 45 days after earnings announcement. To obtain this estimate, we implement the same approaches (5 linear methods and 2 tree-based methods). We estimate these models on a rolling 5 year sample. The only difference is that we add the analyst consensus to the list of possible variables, as well as its squared value and interactions with other variables for our penalized linear estimators.

2 The Term Structure of Forecasting Accuracy

In this section, we compare the accuracy of the three forecasts discussed in Section 1. We will present the mean squared errors of all of our forecasts calculated out-of-sample across the entire sample, after normalizing by the mean squared realized EPS. This normalization is simple to understand, but we give it a quantitative interpretation using the following simple portfolio choice model.

Interpretive model. Assume to simplify that EPS is uncorrelated across firms. The volatility of each firm i 's EPS is assumed to be constant and is denoted σ^2 .⁶ Let γ be the investor's (absolute) risk aversion and θ_i the number of shares bought of firm i . We omit t subscripts here, as we consider the investor solving this problem each year. Consider an investor with utility from her h period ahead forecasts given by:

$$\Pi_h = \sum_i \left(\theta_i EPS_{it+h} - \frac{\gamma}{2} \sigma^2 \theta_i^2 \right)$$

The investor maximizes the expectation of this criterion under her own subjective measure.

⁶It is possible to write down a version of this interpretative model with heterogeneous scaling parameters (variances or price levels). Then, the means would have to be reweighted by the inverse of these scaling parameters. We have experimented with heterogeneous variances sand found little difference with our baseline analysis here.

Let $F_t EPS_{it+h}$ be her subjective forecast of EPS of firm i in year $t+h$. In this case, we can easily derive the loss due to forecasting rule F_t relative to perfect foresight (i.e. knowing the ex-post realization of EPS_{it+h}). It is given by:

$$\frac{\Pi_h^{PF} - \Pi_h^F}{\Pi_h^{PF}} = \frac{MSE_{t+h}^F}{MSS_{t+h}} \quad (3)$$

where Π_h^{PF} is the realized utility under perfect foresight and Π_h^F is the realized utility with the forecasting rule F . MSE_{t+h}^F is the realized mean-squared error of forecasting rule F_t , MSS_{t+h} is the realized mean of squared EPS. Thus, (3) shows normalizing mean-squared errors by the mean squared EPS can be interpreted as the allocative loss compared to a situation of perfect foresight. In Table 1, we report these means for three different horizons in Panels A, B, and C.

Results. The first thing to note from Table 1 is that there are gains from applying supervised machine learning techniques when forecasting EPS, over simpler methods.⁷ The second and third rows of each panel report the MSE (in percent) of two simple approaches. The first one is the random walk predictor: using the most recent realization of EPS as a forecast of next year’s EPS. The second row uses the procedure developed by So (2013), which uses a small number of accounting ratios highlighted by the literature to forecast EPSs and rolling OLS (Fama-MacBeth) regressions. At the one-year horizon, the ML estimators – reported below – have an MSE between 20 and 25% lower than these two simple specifications. This substantial accuracy gain from using machine learning estimators is consistent evidence in Hansen and Thimsen (2020) and van Binsbergen et al. (2020). These findings also justify the use of ML estimators in our subsequent analysis. If we were just using a simple approach, we would underestimate the value added of the machine and therefore overestimate the analyst’s information advantage in Section 3. At longer horizons, there is still a substantial benefit to applying ML techniques, especially tree-based methods.

The second (and more important) result from Table 1 is that there is a *term structure of subjective forecast accuracy*. At the one-year horizon, analyst forecasts are extremely accurate and dominate all variants of our machine forecasts. However, at the two-year horizon, all of our machine forecasts, including the most basic random walk forecast, outperform analyst forecasts. This finding is magnified at the three-year horizon, in which the best machine forecast presents an approximately 40% gain in utility for an investor with the objective function described above.

⁷See Appendix C for the comparison of MSEs across forecasts in different years.

Table 1. The Term Structure of Forecasting Accuracy

This table contains the mean squared error of various forecasts over our entire sample, normalized by the mean squared EPS. Machine and machine + analyst forecasts are formed using the procedure described in Section 1 and Appendix A. The Random Walk forecast is simply earnings-per-share in the current year; the OLS machine forecast is formed following the approach as described in So (2013). The mean squared errors displayed in the table are calculated by first calculating the mean squared error within each year and normalizing by the mean squared realized earnings-per-share in that year, which represents the percentage utility loss relative to having perfect foresight in the interpretive model presented in each year (Section 2). We then take a weighted average across all years with weights proportional to the number of observations in each year (results are insensitive to this weighting) to arrive at the numbers presented in this table, which represent the the average losses to the investor considered in Section 2. Each panel presents results for a different horizon forecast. The final samples in Panels A, B, and C contain and 47,542, 39,973, and 10,831 firm-year observations, respectively.

Panel A: One-Year Horizon Forecasts

	Analyst	Machine	Machine + Analyst
Data	11.04%	-	-
Random Walk	-	20.05%	-
OLS	-	21.86%	-
<i>Penalized linear estimators</i>			
Lasso	-	15.71%	9.65%
Post-Lasso	-	16.22%	9.94%
Iterative Lasso	-	16.36%	9.88%
Ridge	-	17.07%	11.11%
Elastic Net	-	15.72%	9.67%
<i>Tree-based estimators</i>			
Random Forest	-	15.6%	10.03%
Gradient-Boosted Trees	-	15.91%	10.0%

Panel B: Two-Year Horizon Forecasts

	Analyst	Machine	Machine + Analyst
Data	36.67%	-	-
Random Walk	-	34.87%	-
OLS	-	31.73%	-
<i>Penalized linear estimators</i>			
Lasso	-	27.79%	25.21%
Post-Lasso	-	29.2%	26.61%
Iterative Lasso	-	28.94%	26.18%
Ridge	-	28.33%	25.73%
Elastic Net	-	27.79%	25.14%
<i>Tree-based estimators</i>			
Random Forest	-	26.18%	24.2%
Gradient-Boosted Trees	-	27.34%	25.0%

Panel C on Next Page

Table 1. The Term Structure of Forecasting Accuracy (continued)*Panel C: Three-Year Horizon Forecasts*

	Analyst	Machine	Machine + Analyst
Data	42.38%	-	-
Random Walk	-	35.09%	-
OLS	-	30.65%	-
<i>Penalized linear estimators</i>			
Lasso	-	28.24%	26.89%
Post-Lasso	-	30.37%	29.26%
Iterative Lasso	-	29.61%	28.01%
Ridge	-	28.26%	26.97%
Elastic Net	-	28.12%	26.79%
<i>Tree-based estimators</i>			
Random Forest	-	25.06%	24.32%
Gradient-Boosted Trees	-	26.55%	25.93%

The third takeaway from Table 1 is that subjective forecasts are valuable at all horizons. Despite the accuracy of analyst forecasts at the one-year horizon, the machine + analyst forecast improves upon these forecasts by around 10%, which is economically sizeable. This complementarity is still present, but slightly smaller at longer horizons – approximately 5-8%. This improvement from combining subjective and statistical forecasts suggests some complementarity in forecasting – subjective forecasts clearly have information that is not contained in our subjective forecasts, given analyst forecasts are more accurate than machine forecasts at the one-year horizon, but there is evidently some inefficiencies subjective forecasts that statistical techniques can reduce. We formalize this intuition when we discuss noise reduction in Section 3.

Finally, Table 1 shows relatively consistent performance across different supervised machine learning estimators, especially short horizons. At the one-year horizon, estimators that are consistent under (approximate) sparsity conditions (i.e. variants of Lasso and Elastic Net) perform very well, suggesting that the data generating process for earnings may be relatively sparse. However, given that tree-based methods perform relatively well, the strong performance of our sparse estimators may be driven by model uncertainty rather than true sparsity.⁸ The strong performance of Lasso in our setting is similar to evidence on forecasting

⁸Giannone, Lenza, and Primiceri (2018) argue that using the relative performance of methods that assume sparsity to detect sparsity is invalid, since a sparse estimator could arise in sample due to either a truly sparse DGP or large model uncertainty. In the penalized linear estimators we use, one regularization parameter controls both of these, so we cannot distinguish between the two. However, we simply wish to make the point we find less direct evidence *against* sparsity in our setting, relative to the literature that estimates

macro aggregates (Bianchi et al. 2020), but is at odds with from recent evidence in empirical asset pricing suggesting a non-sparse stochastic discount factor (Gu et al. 2018; Kozak et al. 2020; Bryzgalova et al. 2020). So while variation in expected returns are non-sparse, variation earnings seem to be. At longer horizons, sparse estimators continue to perform well, but tree-based methods perform even better, possibly due to increasing complexity in the data generating process.⁹

In sum, the evidence in this section shows subjective forecasts become less accurate relative to statistical forecasts at longer horizons, but contain valuable information orthogonal to the public information in X_{it} at all horizons. In principle, this reduction in relative subjective forecast accuracy could occur for two reasons. First, subjective forecasters could have access to less information orthogonal to X_{it} at longer horizons. For example, analysts may receive strong signals from discussions with management about a firm’s near-term prospects or use high-frequency data sources (Dessaint et al. 2020), but this information may be less valuable for longer horizons. Alternatively, analysts may issue more biased or noisy forecasts at longer horizons, possibly driven by weaker forecast incentives at these horizons or a greater cost of processing public information. In the next section, we perform a decomposition that allows us to separate between these two explanations.

3 Decomposing the Term Structure of Subjective Expectations

In order to structurally interpret the term structure of forecasting accuracy discussed in the previous section, this section decomposes subjective forecasts into three components: information advantage, bias, and noise. Statistical forecasts can reduce noise and bias, but do not have access to all information. In order to isolate and measure each one of these three elements at different forecasting horizons, we write down a simple structural model that we then estimate.

pricing kernels.

⁹Note the tendency to overfit is significantly stronger for tree-based estimators compared to penalized-linear estimators because of the large number of hyperparameters. Lasso-based methods have the benefit of only one parameter (except in the case of Iterative Lasso, in which there is no hyperparametric because the penalty level is theoretically chosen), which makes them less susceptible to overfitting and thus more likely for the results in our sample to be similar on other samples.

3.1 Setup

Denote earnings-per-share at time $t + h$ for firm i as EPS_{it+h} . We consider three time t information sets relevant for forecasting EPS_{it+h} . First, denote the set of publicly available information about EPS_{it+h} observed at time t as X_{it} . In our empirical analysis, this corresponds to the set of variables in Table A1.¹⁰ Secondly, denote the set of information observed by the analyst about EPS_{it+h} at time t as I_{it} . Finally, denote the set of information representing the set of the analyst’s information that is not subsumed by public information as $Z_{it} = I_{it} \setminus X_{it}$. Intuitively, X_{it} represents quantifiable information relevant for forecasting firm-level earnings that is publicly observable, while Z_{it} represents non-quantifiable (soft) or quantifiable private information that is also useful for forecasting firm-level earnings.

Given these information sets, we can decompose EPS_{it+h} (without loss of generality) as

$$EPS_{it+h} = E(EPS_{it+h}|X_{it}) + E(EPS_{it+h}|Z_{it}) + \varepsilon_{it+h} \equiv x_{it}^h + z_{it}^h + \varepsilon_{it+h}, \quad (4)$$

where ε_{it+h} corresponds to the innovation. We assume a constant is included in X_{it} such that $E(Z_{it}) = 0$. Note (4) is an orthogonal decomposition by construction:

$$E(\varepsilon_{it+h}|x_{it}^h, z_{it}^h) = E(z_{it}^h|x_{it}^h) = 0.$$

Let us now discuss how forecasts are formed. In the main text, our core structural assumption is that the consensus analyst forecast is formed as follows:¹¹

$$F_t EPS_{it+h} = g_h(X_{it}) + z_{it}^h + \eta_{it}^h. \quad (5)$$

where $g_h(\cdot) : X_{it} \rightarrow \mathbb{R}$ determines how the analyst maps components of public information into her subjective forecast. Importantly, we place no restrictions on the functional form of

¹⁰There is no reason to believe X_{it} captures all public information about EPS_{it+h} – in principle we could incorporate many other sources of information into it. However, this does not affect the decomposition we do here, in which Z_{it} is *defined* as the information not contained in X_{it} .

¹¹Our choice to model the consensus analyst forecast is driven by the fact that our empirical methodology does not permit identification of analyst-specific bias and noise. In Section 4, we discuss the consequences of this assumption in more detail when we explore models of belief formation.

$g_h(\cdot)$.¹² There are two possible deviations from full-information rationality in this framework:

$$\underbrace{g(X_{it}) \neq x_{it}^h}_{\text{bias}}, \quad \underbrace{\eta_{it}^h \neq 0}_{\text{noise}}.$$

For clarity of exposition, (5) restricts the analyst to be unbiased with respect to private information, but we explore a variant of this model that relaxes this restriction in Appendix D.

Lastly, η_{it}^h is our primary object of interest: the expectation noise term. Broadly speaking, expectation noise could be microfounded in two ways: (i) noisy information models (e.g. Woodford 2003) or (ii) evidence from cognitive psychology (e.g. Khaw et al. 2019; Kahneman et al. 2021).¹³ In the former, noise comes from noise in the agent’s information set. In the latter, noise comes from noise in the internal retrieval and storage of a given set of information (e.g. Woodford 2020). Note that by definition $E(\eta_{it}^h | X_{it}, Z_{it}) = 0$ – the part spanned by X_{it} would be bias not noise.

Within this fairly general framework, we use natural definitions of the forecasters information advantage, bias, and noise. With these definitions, we obtain our key decomposition.

Definition 1. The *information advantage* of the analyst’s forecast at horizon h is defined as $\Theta_h \equiv \text{var}(z_{it}^h)$. The *bias* in the analyst’s forecast at horizon h is defined as $\Delta_h \equiv E[(x_{it}^h - g_h(X_{it}))^2]$. The *noise* in the analyst’s forecast at horizon h is defined as $\Sigma_h = \text{var}(\eta_{t,h})$.

Proposition 1. The difference between the forecasting accuracy of the machine, $MSE_{t+h}^m \equiv E[(EPS_{it+h} - E(EPS_{it+h} | X_{it}))^2]$, and the forecasting accuracy of the subjective forecaster, $MSE_{t+h}^a \equiv E[(EPS_{it+h} - F_t EPS_{it+h})^2]$, can be decomposed as follows:

$$MSE_{t+h}^a - MSE_{t+h}^m = \Delta_h + \Sigma_h - \Theta_h. \quad (6)$$

All derivations are provided in Appendix F. The decomposition in (6) shows the difference in forecasting accuracy between the machine and the analyst mixes three terms: (i) information; (ii) bias; (iii) noise. The informational advantage, Θ_h , represents the value (in terms of MSE) for forecasting EPS of every piece of information in Z_{it} , which is the infor-

¹²We of course assume $g_h(\cdot)$ satisfies regularity conditions such that it can be estimated with the supervised learning techniques we use.

¹³Any elicitation noise or classical measurement error will also be picked up in η_{it}^h . However, we do not emphasize this interpretation because there is little reason to expect these to be large in our setting. Moreover, even if they were, there is no reason why we’d expect them to vary over the forecasting horizon when we use newly updated forecasts.

mation used by the analyst to make her forecast not subsumed by the public information set, X_{it} . This information could be soft (it cannot be put in a dataset), private, or (in our application) it could be public information outside Compustat. Because the machine does not have access to this information, $MSE_{t+h}^a - MSE_{t+h}^m$ is decreasing in Θ_h . Missing on this public information will lead us to overestimate the analyst’s informational advantage.

The other two terms in (6) are the bias and noise. The bias, Δ_h , is restricted to affect public information and may arise for many reasons, such as bounded rationality, cognitive mistakes, or incentives structures that change the analyst’s objective away from minimizing forecast MSE (e.g. Chen and Jiang 2006; Groyberg, Healy, and Maber 2011). We take no stand on the source of this bias. Since machines are unbiased, $MSE_{t+h}^a - MSE_{t+h}^m$ is increasing in the bias. Similarly, this difference is increasing in the noise, Σ_h , which machines are not susceptible to. In Section 4, we explore models in which the bias and noise are endogenous.

The most restrictive assumption in our framework is that the analyst processes the information in Z_{it} without bias. This is made for empirical tractability, as adding a bias on private information makes it more complex to take the model to the data. However, in Appendix D, we discuss the impact of this assumption on our results and show it is possible to extend this framework to allow for and estimate the extent of bias on private information. Specifically, Appendix D extends (5) to include a term in $\alpha_h z_{it}^h$, where $\alpha_h \neq 1$ corresponds to biased treatment of the information in Z_{it} . We estimate α_h by adding two other structural assumptions: (i) Σ_h is inversely proportional to the number of analysts following the firm and (ii) the number of analysts following the firm is independent of Z_{it} . The intuition for the first assumption is each analyst has an independent forecasting noise term, so the noise should decrease in the number of analysts. The second assumption is restrictive and we are not confident in whether it holds, which is why we exclude these results in the main text. We do discuss these results in Section 3.4.

3.2 Moment Selection

In this section we discuss how to identify our three parameters of interest, Θ_h , Δ_h , and Σ_h , using observable moments. The following proposition presents an identification result using three moments that are intuitive.

Proposition 2. *Denote $F_t^* EPS_{it+h}$ and EPS_{it+h}^* as the projections of analyst forecast and*

realized EPS on the space orthogonal to observable information, X_{it} :

$$\begin{aligned} F_t^* EPS_{it+h} &\equiv F_t EPS_{it+h} - E(F_t EPS_{it+h} | X_{it}) = z_{it}^h + \eta_{it}^h \\ EPS_{it+h}^* &\equiv EPS_{it+h} - E(EPS_{it+h} | X_{it}) = z_{it}^h + \varepsilon_{it+h} \end{aligned}$$

Then the information advantage, bias, and noise are identified using the following moments:

$$\begin{aligned} \Delta_h &= E \left[(E(F_t EPS_{it+h} | X_{it}) - E(EPS_{it+h} | X_{it}))^2 \right] \\ \Theta_h &= cov(F_t^* EPS_{it+h}, EPS_{it+h}^*) \\ \Sigma_h &= var(F_t^* EPS_{it+h}) - cov(F_t^* EPS_{it+h}, EPS_{it+h}^*) \end{aligned}$$

Proposition 2 contains straightforward results. The bias is obtained by projecting the forecast error on observables. The information advantage is measured through the covariance between forecasted and realized EPS. The noise is the difference between the variance of forecasts and information advantage. Taken together, these three equations suggest a simple method to estimate the three components in (6) that determine the relative accuracy of statistical and subjective forecasts.¹⁴

3.3 Estimation Procedure

We now discuss estimation of our three parameters of interest using Proposition 2. The approach is straightforward, but we describe it in detail in this sub-section.

First, we flexibly estimate $E(EPS_{it+h} | X_{it})$ and $E(F_t EPS_{it+h} | X_{it})$ using the different supervised machine learning methods described in Section 1.3. Note that $E(EPS_{it+h} | X_{it}) = F_t^m EPS_{it+h}$, which are the machine forecasts we examined in Section 2. For forecasts, we estimate $g_h(X_{it})$ with the same method, replacing EPS_{it+h} with $F_t EPS_{it+h}$ as our dependent variable. This results in one set of firm-level projections, for both EPS and forecasts, per ML technique. By replacing conditional expectations with function approximations from our machine learning estimators, we are implicitly making the assumption that our machine learning estimators are consistent at reasonable rates given our sample size. Without placing further restrictions on the data-generating process for EPS_{it+h} , there are no theoretical results that justify this assumption. However, there is a growing theoretical literature

¹⁴Appendix D shows an identification result in the model extension with bias on the information in Z_{it} , which are more complicated than the simple expressions here.

suggesting this assumption is satisfied under a variety of reasonable DGP assumptions.¹⁵

Secondly, we use these ML estimates to compute, at the level of each firm-year observation (i, t) , the projections of EPS and the forecast onto the variables in X_{it} , as well as EPS_{it+h}^* and $F_t^* EPS_{it+h}$. These capture variation in EPS realizations and analyst forecasts *orthogonal to* X_{it} .

Thirdly, we divide all four projections by the price per share P_{it} . This rescaling is done to address heteroskedasticity issues, as EPS has arbitrary units. We could have directly estimated the ML on normalized EPS and forecasts, but have not done so for comparability of Section 2 with the rest of the literature that directly forecasts the dollar level of EPS (e.g. van Binsbergen et al. 2020).¹⁶

Finally, we use the analogy principle and compute the three sample moments needed to apply Proposition 2 to directly obtain the three structural components. We estimate standard errors of these parameters using a firm-level block bootstrap (we draw entire firm histories to account for within-firm autocorrelation). This standard errors are likely to be too tight because they ignore sampling uncertainty in the predictions generated from our machine learning estimators, effectively treating them as raw data that is resampled directly to compute moments. Here we are constrained by (to our knowledge) the lack of asymptotic results that characterize the behavior of our statistical learning estimators in large samples, so we have no other choice but to take the predictors as given. This approach is standard in the literature (e.g. Patton and Verardo 2012).¹⁷

3.4 Estimation Results

We report the results from estimating our three parameters of interest in Table 2. Each row corresponds to results obtained with one ML approach. To ease comparability with Table 1, we normalize the three components by the mean sum of square of EPS normalized by prices: $\frac{1}{N} \sum_{it} \left(\frac{EPS_{it+h}}{P_{it}} \right)^2$. The interpretation is the same Section 2 – each one of the components is

¹⁵For asymptotic results on the approximation error of various supervised learning estimators under different DGP assumptions in large samples, see Belloni et al. (2011) for Iterative Lasso, Chetverikov, Liao, and Chernozhukov (2020) for cross-validated Lasso, Wager and Athey (2018) for random forests, and Schmidt-Hieber (2020) for deep neural networks.

¹⁶This normalization is not crucial to our findings, but significantly improves the distribution properties of the data.

¹⁷We could in principle bootstrap the data and re-estimate our models, but this is too computationally intensive. Moreover, it's not clear that every estimator satisfies the regularity conditions required for the bootstrap to be asymptotically valid (Horowitz 2001).

a contribution to the loss of profits relative to the perfect foresight rule, but now we adjust for the price normalization discussed in the prior section.

Table 2. The Term Structure of Information, Bias, and Noise

This table presents our results from estimating the three parameters in our model using the estimation procedure described in Section 3.3. As described in Section 3.1, Θ_h represents the analyst information advantage; Δ_h represents the analyst bias; Σ_h is the analyst noise. Each row corresponds to the use of a different ML estimator to estimate the two conditional expectation functions described in Section 3.3. Each panel focuses on a different horizon forecast. Parameter estimates are normalized by the average squared earnings-per-share divided by price, calculated across the entire sample. Standard errors are calculated using a clustered bootstrap at the firm level with 1,000 iterations. For each estimation, the firm-year level variables in Proposition 2 are trimmed at 5 times the interquartile range, after they are scaled by price. See Section 3.3 for additional details on our estimation procedure. The final samples in Panels A, B, and C contain and 41,208, 35,458, and 11,533 firm-year observations, respectively.

Panel A: One-Year Horizon Forecasts

Estimator	Θ	s.e.	Δ	s.e.	Σ	s.e.
Lasso	5.58%	(0.13%)	2.12%	(0.03%)	1.46%	(0.07%)
Post-Lasso	6.26%	(0.13%)	3.31%	(0.04%)	1.9%	(0.08%)
Iterative Lasso	6.24%	(0.12%)	1.56%	(0.02%)	1.5%	(0.09%)
Ridge	7.0%	(0.14%)	3.27%	(0.04%)	2.16%	(0.1%)
Elastic Net	5.63%	(0.12%)	2.1%	(0.03%)	1.43%	(0.09%)
Random Forest	4.31%	(0.09%)	1.99%	(0.03%)	1.17%	(0.07%)
Gradient-Boosted Trees	4.73%	(0.11%)	2.36%	(0.03%)	1.45%	(0.07%)
Mean	5.68%		2.39%		1.58%	

Panel B: Two-Year Horizon Forecasts

Estimator	Θ	s.e.	Δ	s.e.	Σ	s.e.
Lasso	4.6%	(0.15%)	10.62%	(0.14%)	4.85%	(0.17%)
Post-Lasso	5.05%	(0.22%)	13.53%	(0.17%)	6.02%	(0.2%)
Iterative Lasso	5.86%	(0.24%)	8.73%	(0.11%)	5.71%	(0.2%)
Ridge	5.52%	(0.2%)	12.3%	(0.17%)	6.31%	(0.18%)
Elastic Net	4.63%	(0.18%)	10.65%	(0.16%)	4.82%	(0.17%)
Random Forest	3.11%	(0.14%)	11.66%	(0.16%)	4.4%	(0.16%)
Gradient-Boosted Trees	3.65%	(0.17%)	12.0%	(0.18%)	4.23%	(0.18%)
Mean	4.63%		11.36%		5.19%	

Panel C: Three-Year Horizon Forecasts

Estimator	Θ	s.e.	Δ	s.e.	Σ	s.e.
Lasso	5.51%	(0.36%)	11.46%	(0.22%)	8.41%	(0.43%)
Post-Lasso	6.33%	(0.4%)	16.97%	(0.37%)	9.57%	(0.52%)
Iterative Lasso	8.17%	(0.45%)	8.4%	(0.14%)	8.8%	(0.49%)
Ridge	5.51%	(0.36%)	13.52%	(0.28%)	10.22%	(0.41%)
Elastic Net	5.69%	(0.43%)	11.43%	(0.22%)	8.09%	(0.46%)
Random Forest	4.15%	(0.34%)	11.54%	(0.25%)	7.1%	(0.33%)
Gradient-Boosted Trees	4.0%	(0.31%)	14.82%	(0.33%)	8.38%	(0.38%)
Mean	5.62%		12.59%		8.65%	

Panel A presents the results for one-year ahead forecasts. Here the analyst information advantage, Θ_1 , is larger than the bias and the noise combined, $\Delta_1 + \Sigma_1$. Given (6), this result is consistent with the fact that $MSE_{t+1}^a < MSE_{t+1}^m$ in Panel A of Table 1 – analysts outperform at short horizons because they have a significant information advantage. Although the bias and noise are smaller than the information advantage at this horizon, they are non-negligible. This is consistent with the gain to combining subjective and statistical forecasts in Panel A of Table 1.

Panels B and C examine the information, bias, and noise for two-year and three-year ahead forecasts. The first thing to note from these panels is the information advantage does *not* decay with the horizon. According to (6), this implies the decline in subjective forecast accuracy across horizons in Table 1 must be driven and increase in bias and/or noise. This is exactly what we find – noise increases approximately linearly across horizons, while the bias jumps between between the one and two-year horizons and increases slightly between the second and third-year horizons. This non-linearity in the increase in bias is driven by sample composition, as discussed below. These estimation results in Table 2 appear relatively stable across the different supervised machine learning estimators.

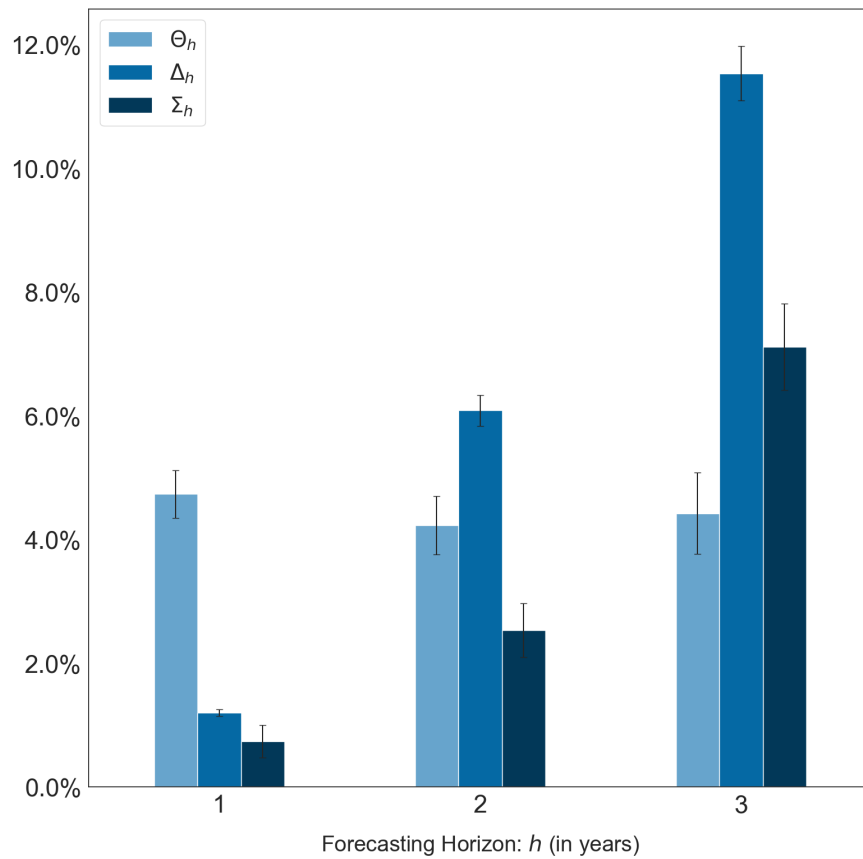
The results in Table 2 are difficult to interpret because moving from one-year to three-year horizon forecasts changes the sample composition significantly — most firms in I/B/E/S have data on one-year ahead forecast, but only larger, older, and more profitable firms have data on three-year ahead forecasts. Notably, the number of observations only decreases slightly between one-year and two-year ahead forecasts, while it decreases by a factor of four at the three-year horizon.

In Figure 1, we plot the term structure of our three parameters estimated on the sub-sample of firm-year observations for which we observe a consensus analyst forecast at *all three* horizons. The results show that the term structure of the analyst information advantage is still essentially flat, and the term structure of bias and noise are still upward sloping. However, the term structure of bias and noise are now increasing approximately linearly.

As discussed above, we also explore a variant of this framework that allows for bias in private information, but requires strong assumptions. The results in Appendix D show the term structure of Δ_h is unaffected. We also find evidence of an upward sloping term structure of bias on the information in Z_{it} , α_h , with $\alpha_h \approx 1$ at short-horizons and $\alpha_h > 1$ at long horizons. We continue to find an upward sloping term structure of noise, albeit less steep, and the term structure of Θ_h becomes slightly downward sloping.

Figure 1. The Term Structure of Information, Bias, and Noise

This table presents our results from estimating the three parameters in our model using the estimation procedure described in Section 3.3 on the sub-sample of firm-year observations for which we have analyst forecast data at all three forecasting horizons. As described in Section 3.1, Θ_h represents the analyst information advantage; Δ_h represents the analyst bias; Σ_h is the analyst noise. For the results plotted here, we used Elastic Net to estimate the two conditional expectation functions. Parameter estimates are normalized by the average squared earnings-per-share divided by price, calculated across this entire sub-sample. For each estimation, the firm-year level variables in Proposition 2 are trimmed at 5 times the interquartile range, after they are scaled by price. See Section 3.3 for additional details on our estimation procedure. This sample contains 10,831 firm-year observations. Black bars represent 95% confidence intervals based on a clustered bootstrap at the firm-level with 100 iterations.



In sum, the evidence in Table 2 and Figure 1 suggest there is a strong upward sloping term structure of bias and noise, but a relatively flat term structure of the analyst information advantage. In the next section, we examine the model ingredients needed to *jointly* match this term structure of bias and noise. Henceforth, we ignore the term structure of information since it is flat, although there is likely heterogeneity in the cross-section of firms that could be explored.

3.5 Mitigating Forecast Noise and Bias

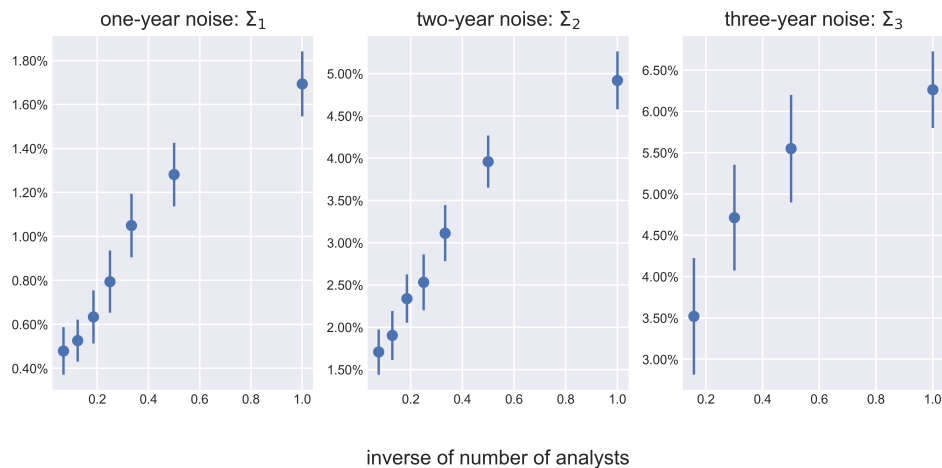
In this last sub-section, we explore two options to mitigate forecasting noise and bias. The first option, extensively discussed in Kahneman et al. (2021), consists of using a fixed rule to discipline subjective forecasts. The intuition for the effectiveness of this strategy for reducing noise follows from standard state-space models (e.g. Kalman filter). In these models, the optimal way to use a noisy signal is to apply some shrinkage. This is essentially what the machine + analyst forecast formed in Section 1.4 implements. Column 4 of Table 1 shows the performance of this forecast, using various techniques to optimally combine the data available to the econometrician and subjective forecasts. The relative “weights” (these methods are non-linear) of the two types of information are chosen optimally using the various ML techniques described above. At the one-year horizon, the improvement brought by these optimal rules is modest (about 10% of MSE), but at longer horizons, the improvement becomes much larger (about 33% reduction in MSE). This is because long-term subjective forecasts are both strongly noisy and biased.

The second option to tame noise directly is to increase the number of forecasters. To illustrate this, we generate estimates of noise at $h = 1, 2, 3$ across bins of firm-years based on the number of analysts that make up the consensus forecast at horizon h , denoted N_{it}^h . The estimation strategy we use to do so is intuitive and requires one meaningful identifying assumption discussed in Appendix B. In Figure 2, we report estimated the level of noise on the y-axis and $\frac{1}{N_{it}^h}$ on the x-axis. We expect the relationship to be linear, with the intercept term being equal to zero if noise is perfectly uncorrelated across analysts. Looking at the Figure 2, several patterns emerge. First, the relationship between noise and inverse number of analysts is approximately linear at all three horizons. Secondly, the reduction of noise from going from 1 to 10 analysts is sizable: at the one-year horizon, it buys 1.2ppt MSE reduction of MSS (not huge because noise is small at this horizon). At three-year horizon, it buys a more significant 3.5ppt reduction. Finally, even with an infinite number of forecasters, noise

does not seem to disappear, which suggesting correlation in noise – we estimate it would go from 1.8% to 0.4% at one-year horizon and from 6.5% to 3.2% at three-year horizon.

Figure 2. Noise and the Number of Forecasters

This figure plots our estimates of noise at three different horizons across evenly-spaced bins of $\frac{1}{N_{it}^h}$, where N_{it}^h is the number of analysts that make up the consensus forecast firm i in year t at horizon h . We form moments using the Random Forest for statistical forecasts. To estimate conditional noise, we perform our conditional estimation described in Section 3.5 and Appendix B. Prior to conducting the estimation, the firm-year level variables in Proposition B1 are trimmed at 5 times the interquartile range, after they are scaled by price. Estimates are normalized by the mean squared EPS across the entire sample. Confidence intervals are calculated using a clustered bootstrap at the firm level with 1,000 iterations and have 90% joint coverage across all points within each figure. Critical values for these joint confidence intervals are calculated using the 1,000 bootstrap iterations. The final sample used to produce these figures has 47,542, 39,973, and 13,066 firm-year observations at the one-year, two-year, and three-year horizons respectively. For noise at the three-year horizon, there are only four bins because there is less variation in the number of analysts.



4 Models of the Term Structure of Bias and Noise

This section explores the extent to which existing models can *jointly* fit our estimated patterns in the term structure of bias and noise. We first explore a large class of existing models that deviate from full-information rational expectations. We show these models, in their simplest form, cannot jointly match the sign and magnitude of the slopes of these term structures. Thus, a mechanism needs to be added – we propose a simple model based on a noisy cognitive default. This mechanism is portable and could be added to existing models.

4.1 Existing Models

We first consider a list of existing models for the term structure of bias and noise. We primarily focus on variants and extensions of noisy information models, which is a standard

framework that has predictions on the term structure of expectation noise and bias. This noise is crucial to our setting because we can clearly reject the null hypothesis of noiseless forecasts. Thus, we will not consider some popular used models of belief formation, such as backward looking models (e.g. adaptive expectations and pure extrapolation).

Setup. Recall the framework described in Section 3.1, where we have

$$\begin{aligned} EPS_{it+h} &= x_{it}^h + z_{it}^h + \varepsilon_{it+h}, \\ F_t EPS_{it+h} &= g_h(X_{it}) + \eta_{it}^h + z_{it}^h. \end{aligned} \tag{7}$$

As discussed previously, the first equation is a normalization and the second equation is an assumption. Each model we consider delivers a forecasting equation of the form in (7), where $g(X_{it})$ and the variance of η_{it} are determined by the primitives of the model. We continue to assume the analyst is unbiased with respect to private information.

To clarify the discussion, we now impose additional structure on the data-generating process to be able to analytically compare multiple existing models (which work with AR1 processes). Importantly, our conclusion in this section that the law of iterated expectations embedded in most models makes it difficult to fit our evidence does not rely on this assumption.

Assumption 1. Denote $x_{it} \equiv x_{it}^1$. The data generating process for x_{it} is:

$$x_{it} = \mu(1 - \rho) + \rho x_{it-1} + u_{it}, \quad E(u_{it} | x_{it-1}) = 0, \quad \text{var}(x_{it}) \equiv \sigma_x^2.$$

Given we have estimated $E(EPS_{it+h} | X_{it}) = F_t^m EPS_{it+h}$, we can test the validity of this assumption. Table A8 provides evidence that an AR(1) is a reasonable description of the data.

4.1.1 Baseline Noisy Information Model

We first consider a baseline noisy information model in the spirit of Woodford (2003). This noisy information framework is the most natural starting place, given it is a classic model that generates both bias (from the viewpoint of the econometrician) and noise. In this model, the analyst observes noisy signals of x_{it} , denoted \mathcal{S}_{it} , but is an otherwise rational (i.e. Bayesian) forecaster that knows μ and ρ . Given \mathcal{S}_{it} , the analyst applies Bayes rule to

form forecasts as follows:

$$F_t x_{it} = E(x_{it} | \mathcal{S}_{it}), \quad F_t x_{it+h} = (1 - \rho^{h-1})\mu + \rho^h F_t x_{it}. \quad (8)$$

In this setting, an explicit characterization of the analyst's forecasts requires placing structure on \mathcal{S}_{it} and the innovation DGP u_{it} . For example, assuming the analyst observes $s_{it} = x_{it} + v_{it}$ each period and v_{it} is normally distributed results in the Kalman filtering solution, where $F_t x_{it} = G_{it} s_{it} + (1 - G_{it}) F_{t-1} x_{it}$ and G_{it} is the Kalman gain. However, regardless of the structure of \mathcal{S}_{it} , we can characterize the slope of the term structure of bias and noise implied by this model.

Proposition 3. *In the baseline noisy information model, the term structure of bias and noise are downward sloping with the same slope:*

$$\forall h > 1 : \frac{\Delta_h}{\Delta_1} = \frac{\Sigma_h}{\Sigma_1} = \rho^{2(h-1)} < 1.$$

In this model, the intuition for the downward sloping term structure of bias and noise follows from considering the extreme case in which $h \rightarrow \infty$. Because x_{it} is a stationary process, the best infinite horizon forecast is the long-run mean, μ , which the analyst knows. Thus, the analyst will be unbiased in this extreme case and will also issue noiseless forecasts because the analyst will place no weight on her sequence of noisy signals. As h shrinks towards 1 from this extreme case, the analyst relies more on noisy signals and less on her knowledge of the long-run mean. This introduces bias and noise in her forecast, resulting in downward sloping term structures. The evidence in Section 3.4 provides a clear rejection of the baseline noisy information model, as we find upward sloping term structures of bias and noise.

4.1.2 Variants of Baseline Noisy Information Model

The baseline noisy information model of the previous section is characterized by the following two equations:

$$F_t x_{it} = E(x_{it} | \mathcal{S}_{it}), \quad (9)$$

$$F_t x_{it+h} = (1 - \rho^{h-1})\mu + \rho^{h-1} F_t x_{it}. \quad (10)$$

Prior literature has proposed multiple variants of this baseline model than can be seen as simply adjusting one of these two equations. We now discuss the predictions of commonly-used variants one-by-one.

Bounded rationality. A common microfoundation for noisy information models is bounded rationality (e.g. Sims 2003). In these models, the set of signals, \mathcal{S}_{it} , is endogenously chosen to maximize an objective function decreasing in forecast errors, subject to a cost function increasing in the mutual information of the signals, \mathcal{S}_{it} , and the state variable, x_{it} . Thus, these models amount to changing the form of \mathcal{S}_{it} in (9). Although these models introduce a tight connection between the signals and primitives (e.g. signal informativeness and cognitive capacity), they also have downward sloping term structures of bias and noise. This follows from the fact that Proposition 3 was derived without requiring placing structure on the distribution on \mathcal{S}_{it} .

Diagnostic expectations. Bordalo et al. (2020) combine diagnostic expectations, which generates overreaction to recent news, with noisy information. This model breaks (9) due to non-rational expectations. In its simplest form, this model exhibits a downward sloping term structure of bias and noise for the same reason as the baseline noisy information model: the analyst knows x_{it} is mean-reverting, so she shrink her forecasts towards the long-run mean, μ , as h grows. When the DGP is an AR1 process, as in assumption 3 (and in most applications of this model), bias and noise monotonically decrease with horizon.

Over-confidence. Another common way to generate non-rational reactions in the noise information framework is via agents' overconfidence about their signal qualities (e.g. Daniel et al. 1998; Eyster, Rabin, and Vayanos 2019). In the common case where \mathcal{S}_{it} consists of signals each period with independent normal errors, over-confidence corresponds to the analyst updating using a variance lower than the true signal variance. Since overconfidence only changes (9), it will only change the level of bias and noise, but both term structures will still be downward sloping.

In sum, noisy information models and their commonly-used variants need an additional mechanism in order to match the term structure of bias and noise we document. As they are currently written, the forecaster knows μ and ρ and forecasts the term structure according to (10). Thus, the law of iterated expectations holds, implying the forecaster cannot be biased or noisy as $h \rightarrow \infty$.

4.1.3 Alternative Frameworks that Break Condition (10)

The previous section shows noisy information models cannot match the term structure of bias and noise because (10) holds. We now consider three sets of models that break this equation in different ways.

Unknown persistence. The first way to break (10) is to assume the forecaster believes the persistence of x_{it} is $\hat{\rho} > \rho$, while maintaining noisy information. Angeletos et al. (2020) show this over-extrapolation is necessary for matching over-reaction in macroeconomic expectations. Although this type of over-extrapolation does make the term structure of bias and noise less downward sloping in our setting, the only way for this to deliver upward sloping term structures is to assume $\hat{\rho} > 1$, which Table A8 provides evidence against.

Unknown mean. The second way to break (10) is to assume the forecaster believes the long-run mean of x_{it} is $\hat{\mu} \neq \mu$. Afrouzi et al. (2021) provide a model of this sort, where the forecaster does not know μ and consequently overweights recent information in the estimation of it. In Appendix E, we show this model has the potential to qualitatively match our data, as long-run forecasts are further from the rational expectation, but cannot do so quantitatively.

Failure of law of iterated expectations. Fuster, Laibson, and Mendel (2010) propose a framework in which the law of iterated expectations fails, known as natural expectations. In this framework, the true DGP is a stationary AR(2) in levels, but the forecaster has an “intuitive” DGP that is an AR(1) in changes. Importantly, because the intuitive model is non-stationary, this model gets a larger weight at longer horizons, which generates an upward sloping term structure of bias. This model does not have a source of noise, but the framework we introduce next is motivated by this idea that the failure of the law of iterated expectations can generate long-run forecasts further away from the rational expectation.

4.2 Proposed Model

Setup. We next present a model with a mechanism that can qualitatively and quantitatively match the upward sloping term structures of bias and noise. In our model, the forecaster issues forecasts according to:

$$F_t[E(EP S_{it+h}|X_{it})] = (1 - m_h)x_{it}^d + m_h E(EP S_{it+h}|X_{it}), \quad (11)$$

$$= (1 - m_h)x_{it}^d + m_h x_{it}^h = (1 - m_h)x_{it}^d + m_h F_t^m EPS_{it+h}.$$

This forecasting equation is motivated by the limited attention framework of [Gabaix \(2014\)](#), where x_{it}^d corresponds to a cognitive default and $m_h \in [0, 1]$ represents the horizon-specific amount of attention paid to processing the public information contained in X_{it} . If $m_h = 1$, the analyst issues forecast equal to the conditional expectation (i.e. the machine's forecast). Unlike [Gabaix \(2014\)](#), we assume x_{it}^d is a random variable, which we parameterize as follows:

$$x_{it}^d = \alpha + \beta EPS_{it} + v_{it}, \quad E(v_{it} | X_{it}, Z_{it}) = E(v_{it}) = 0, \quad \text{var}(v_{it}) \equiv \sigma_v^2.$$

The default, x_{it}^d , may have a fixed bias, α , is partly anchored on the current value of the state variable, β (in the spirit of the availability heuristic), but also contains noise, v_{it} .¹⁸ The fact that x_{it}^d is a *noisy default* is crucial to this framework, which could be micro-founded using evidence in cognitive psychology that suggests individuals' decision-making has significant variation orthogonal to observables (e.g. [Azeredo da Silveira et al. 2020](#); [Enke and Graeber 2020](#); [Kahneman et al. 2021](#)).

To discipline the term structure of m_h , we follow [Patton and Timmermann \(2010\)](#) and assume:

$$m_h = \frac{\kappa^2}{\kappa^2 + MSE_{t+h}^{FIRE}}, \quad (12)$$

where

$$MSE_{t+h}^{FIRE} = MSE_{t+h}^m - \Theta_h = E[(EPS_{it+h} - E(EPS_{it+h} | X_{it}, Z_{it}))^2]$$

is the MSE of the full-information rational expectations (FIRE) forecast. This specification of m_h captures of a form of bounded rationality: as the rational forecast error increases, the forecaster relies more on her default.¹⁹ The extent to which the rational forecast error matters is captured by κ . As $\kappa \rightarrow 0$, the forecaster always relies on her default. As $\kappa \rightarrow \infty$, the forecaster becomes perfectly rational.

Term structure of bias and noise. Within our proposed model, we can calculate the term structure of bias and noise.

¹⁸Our model will fit the term structure of bias and noise even if we restrict $\alpha = \beta = 0$. We choose to allow for an arbitrary α and β so that we can target the $EF_t EPS_{it+h}$ as one of our moments in the estimation.

¹⁹In principle we could formula m_h as being chosen by the forecaster based on a bounded rationality problem in the spirit of [Sims \(2003\)](#). In that case, κ would be an endogenous object, depending on information constraint faced by the forecaster. We have chosen not to do this for simplicity in estimation (following [Patton and Timmermann 2010](#)), as any of these models would generate a similar formula for m_h , with the difference being the specific microfoundation for κ . In all cases, as the forecasters bounded rationality constraint is relaxed, we would obtain $\kappa \rightarrow \infty$.

Proposition 4. *The term structure of bias and noise are given by:*

$$\Delta_h = (1 - m_h)^2 \left[E\left((x_{it}^h)^2\right) + \alpha^2 + \beta^2 E\left(EP S_{it}^2\right) + 2\alpha\beta E\left(EP S_{it}\right) - \right. \\ \left. 2\alpha E\left(x_{it}^h\right) - 2\beta E\left(EP S_{it}x_{it}^h\right) \right], \\ \Sigma_h = (1 - m_h)^2 \sigma_v^2.$$

The formula for noise is simple: all expectation noise comes from noise in the prior, so noise increases as the weight on the prior increases (i.e. m_h decreases). To build intuition for the formula for Δ_h , consider the case where $\alpha = \beta = 0$. Then $\Delta_h = (1 - m_h)^2 E(x_{it}^h)^2$. This expression shows bias increases as the weight on the prior increases and further depends on the second moment of the conditional expectation, x_{it}^h .

These two expressions illustrate how our model delivers a tight link between bias and noise through bounded rationality. Therefore, in order to match the upward sloping term structure of bias and noise in the data, we need m_h to decrease as h increases. Given (12), this is the case because Table 1 shows MSE_{t+h}^m increases with the horizon, while Table 2 shows Θ_h is approximately constant. These expressions also show that a model in which only σ_v varies across horizons will not work because it will not generate a term structure of bias.

In sum, our model is parametrically parsimonious with three key parameters: (i) σ_v^2 , which pins down by the average level of noise in subjective forecasts; (ii) α , which pins down the average level of optimism; (iii) κ , which determines the relative weight of the rational forecast and governs the term structure of both noise and bias. The parameter β is less critical and is only here to ensure forecasts line up with EPS levels in the cross-section.

Estimation. We estimate our model using Generalized Method of Moments (GMM). Our parameter vector is

$$\theta = \left(\kappa \quad \alpha \quad \beta \quad \sigma_v \right)^T \in \mathbb{R}^4.$$

We target the following nine moments: bias at all three horizons, noise at all three horizons, and the mean analyst forecast at all three horizons. We then perform an iterative GMM²⁰ to generate our estimate of the parameter vector. To calculate standard errors, we perform

²⁰Specifically, we solve the GMM objective using an identity weighting matrix, generate an estimate of the optimal weighting matrix, then solve the GMM objective again. We continue this iteration, solving the objective and updating the weighting matrix, until each element of the parameter vector changes by less than 1%. This takes between 2 and 3 iterations depending on the bootstrap sample.

a clustered bootstrap at the firm-level, and then re-estimate the iterative GMM on each bootstrap sample.²¹

Table 3. Model Estimation Results

This table presents the results from estimating the model in Section 4.2 using an Iterative GMM. The model has four parameters and we target nine moments – the term structure of bias and noise, as well as the mean analyst forecast at all three horizons. The sample used here is the same sample as in Figure 1. To calculate bias and noise, we use Elastic Net to estimate the two conditional expectation functions. Standard errors are based on a clustered bootstrap at the firm-level with 100 iterations.

	κ	α	β	σ_v
GMM Estimate	0.0709	0.0191	2.648	0.0774
Std. Error	(0.0017)	(0.0019)	(0.0783)	(0.0034)
Implied m_h	$m_1 = 0.95$	$m_2 = 0.85$	$m_3 = 0.78$	

Results. The results from our GMM estimation are presented in Table 3. We first note that our standard errors are relatively small for all four model parameters, suggesting the model is robust to local misspecification (Andrews, Gentzkow, and Shapiro 2017). Turning to the parameter estimates, we find evidence of significant noise in the default: $\sigma_v \approx 0.08$. For reference, $\text{sd}(x_{it}^1) \approx 0.037$, implying the noisy we estimate in the default is around two times as large as the noise in the data generating process. This significant cognitive noise is needed to match the average level of noise across the three horizons and we discuss it further in Section 5.1. We also find $\alpha > 0$ and $\beta > 0$, suggesting there is some fixed optimism bias in the prior of analysts and a tendency to over-weight the current observation, in the spirit of the availability heuristic. The combination of the fixed bias and availability heuristic help match the average level of bias across the three horizons, but are not crucial to match the slope of the bias (or noise) term structure.

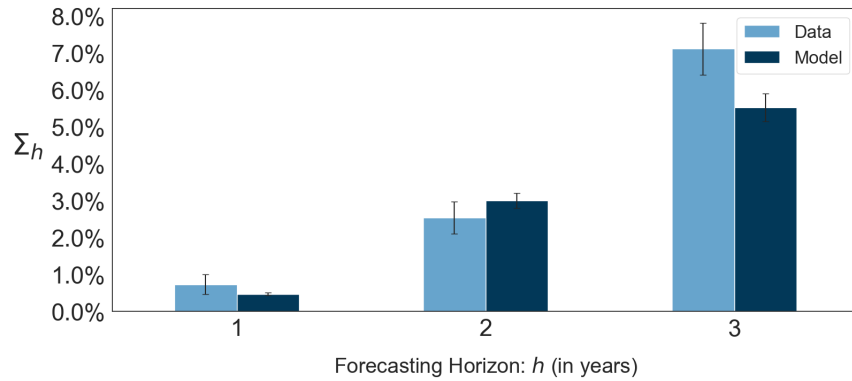
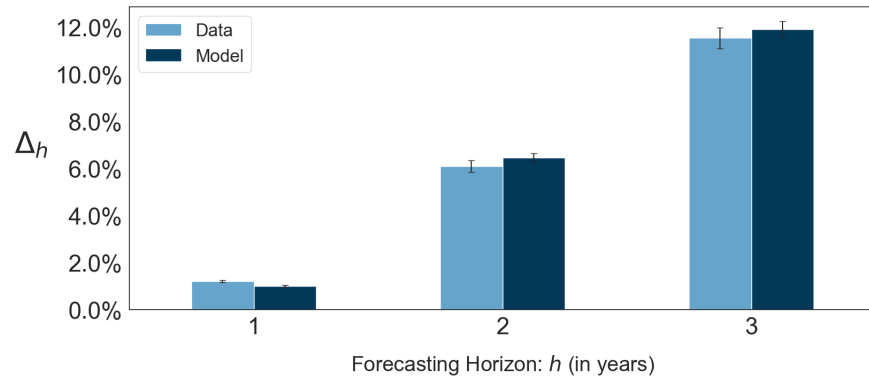
Our model’s most important parameter is κ , which determines the *slope* of the term structure of bias and noise through m_h . We estimate $\kappa \approx 0.07$. To interpret this estimate, Table 3 presents the implied values of m_h from this value of κ , which are $m_1 = 0.95$, $m_2 = 0.85$, and $m_3 = 0.78$. This declining term structure of m_h is crucial for the ability of our model to fit the data, as an upward sloping term structure of bias and noise in our model comes from forecasters relying more on their biased and noisy defaults at longer horizons.

Figure 3 plots the term structure of bias and noise in our estimated model versus the

²¹We use the bootstrap procedure from Hahn (1996), in which the GMM objective function is not re-centered before estimation on each bootstrap sample. This generates standard errors with valid coverage (asymptotically) even under model misspecification.

Figure 3. The Term Structure of Information, Bias, and Noise: Model and Data

This table presents a comparison of the term structures of bias and noise in the data versus in the model in Section 4.2. The moments in the model are calculated using Proposition 4 and the parameter estimates in Table 3. The sample used here is the same sample as in Figure 1. To calculate bias and noise, we use Elastic Net to estimate the two conditional expectation functions. Black bars represent 95% confidence intervals based on a clustered bootstrap at the firm-level with 100 iterations.



data. The results show our model reproduces the term structure of bias and noise very well, due to the decreasing term structure of m_h in Table 3. The moments in our model are not statistically different than the moments in the data, except for noise at the three-year horizon. The inability to fit Σ_3 is partly driven by the fact that we use the GMM optimal weighting matrix and Σ_3 is the least precisely estimated. Nevertheless, we note the term structure for noise looks more convex than the term structure for bias, which is something that could be explored in more detail if data on longer horizon forecasts were available.

Comparison of estimated m_h with existing literature. The key equation in our model, (11), is motivated by the limited attention framework of Gabaix (2014). Because of the similarities between these frameworks, we can compare the values of m_h implied by our estimate of κ to literature that estimates the attention parameters in the Gabaix (2014) framework. In Gabaix (2014), an agent perceives the value of a state variable, x , to be $x^s = (1 - m)x^d + mx$, where $m \in [0, 1]$ represents the agent’s attention paid to the state variable and x^d is the agent’s cognitive default. Gabaix (2019) reviews the literature that estimates m and finds a strong positive relationship across studies between the estimated m and a measure of attribute importance (AI), which captures how much using $x^s = x$ versus $x^s = x_d$ matters for the agent’s decision utility. This positive relationship is consistent with a form of bounded rationality.

In our setting, the analog of AI from Gabaix (2019) at horizon h is $\frac{\text{sd}(x_{it}^h)}{\text{sd}(EPS_{it})}$, which represents the ratio of the standard deviation of the true state variable to that of the forecaster’s default. This ratio is 0.85, 0.82, and 0.81 at $h = 1, 2, 3$, respectively. These values are high compared to Allcott and Wozny (2014), which has an AI²² of 0.58 and estimate $m = 0.76$. Since Allcott and Wozny (2014) have the highest AI of the studied surveyed by Gabaix (2019), we would expect their estimated $m = 0.76$ to be a lower bound on our estimates of m_h . This is indeed the case, suggesting analysts are relatively attentive in our setting because of large differences in forecasting performance between the random walk forecasting rule, $F_t EPS_{it+h} = EPS_{it}$, and the true conditional expectation, $F_t EPS_{it+h} = x_{it}^h = F_t^m EPS_{it+h}$.²³

²²See Gabaix (2019), Table 1, for a calculation of AI across different studies.

²³Gabaix (2019) formally calibrates an attention function, $m = A(\text{AI})$, which maps attribute importance to attention. Using non-linear least squares, he estimates $A(x) = \max\left(\frac{\frac{x}{0.09} - 1}{\frac{x}{0.09} + 44.8}, 0\right)$ using data on estimated m ’s and AIs across studies. Applying this function to our setting gives $m_1 \approx 0.95$ and $m_2 \approx m_3 \approx 0.94$. This value of m_1 is close to what we estimate, but the values of m_2 and m_3 are too high, suggesting more curvature in the attention function for high values of AI is needed.

5 Additional Model Predictions

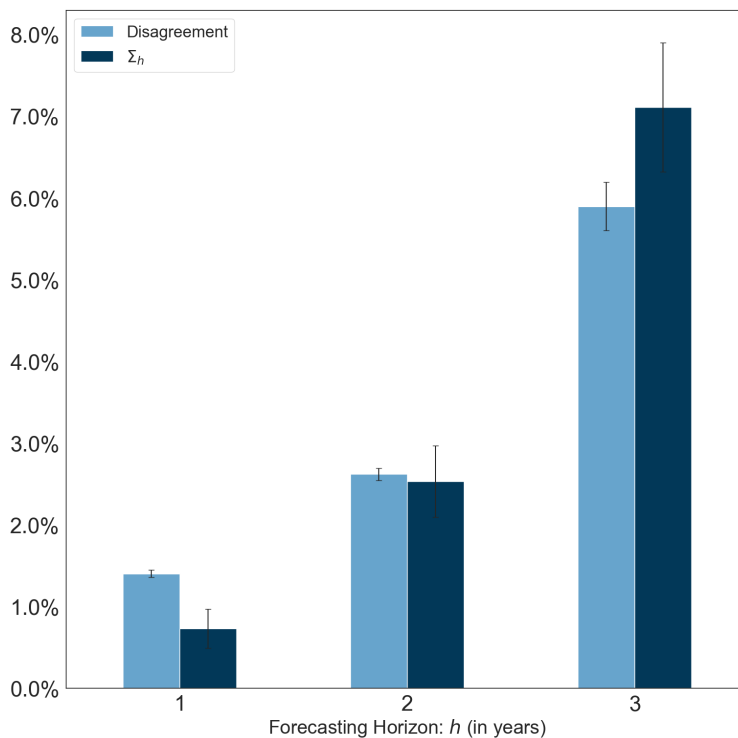
In this section, we discuss how the model proposed in Section 4.2 can quantitatively match two non-targeted moments: (i) an upward sloping term structure of disagreement and (ii) a positive relation between noise and volatility.

5.1 Term Structure of Disagreement

The existing models discussed in Section 4.1 that fail to match the upward sloping term structure of bias and noise also predict downward sloping term structures of disagreement, a point made by [Patton and Timmermann \(2010\)](#). This happens because forecasters know the parameters of the true data generating process and shrink their forecasts to the long-run mean as the forecasting horizon increases, reducing disagreement across forecasters.

Figure 4. The Term Structure of Forecaster Disagreement and Noise

This table presents a comparison of the term structures of forecaster disagreement and noise in the data. To calculate forecaster disagreement, we first calculate the variance across analyst forecasts of horizon h for a given firm-year. As in Section 3.3, we next normalize this variance by price-per-share squared. We then average this variance across all-firm years for each horizon h , using the same sample of firm-year observations in Figure 1. To calculate noise, we use Elastic Net to estimate the two conditional expectation functions. Finally, we normalize both the noise and disagreement by the average squared earnings-per-share divided by price, calculated across the entire sample (Section 2). Black bars represent 95% confidence intervals based on a clustered bootstrap at the firm-level with 100 iterations.



In Figure 4, we plot the term structure of forecaster disagreement, calculated as variance of the individual analyst forecasts that make up the consensus forecast averaged across firm-years. Contrary to the predictions of the models in Section 4.1, the results show the term structure of disagreement is upward sloping. Figure 4 also shows the values of disagreement and noise are similar at all three horizons. This is exactly what our model would predict: in our model all disagreement across forecasters comes from heterogeneity in defaults due to v_{it} . Thus, another way to interpret our estimate of σ_v is as a measure of prior heterogeneity.

Patton and Timmermann (2010) similarly document an upward sloping term structure of disagreement among professional private sector GDP and inflation forecasters. Using a similar framework to the one in Section 4.2, they estimate the standard deviation of priors across forecasters is 2.5 times as large as the standard deviation of the forecasting variable for GDP and 5 times as large for inflation.

Since the standard deviation of priors across forecasters is equal to σ_v in our model, we can similarly calculate the standard deviation of forecaster priors relative to the forecasting variable, $\frac{\sigma_v}{\text{sd}(x_{it}^1)}$. Using the estimate of $\sigma_v \approx 0.08$ from Table 3 and $\text{sd}(x_{it}^1) \approx 0.037$, we find $\frac{\sigma_v}{\text{sd}(x_{it}^1)} \approx 2.2$. Thus, we get a quantitatively similar estimate of heterogeneity in forecasters' priors to Patton and Timmermann (2010) in a very different forecasting setting.

5.2 Noise and Earnings Volatility

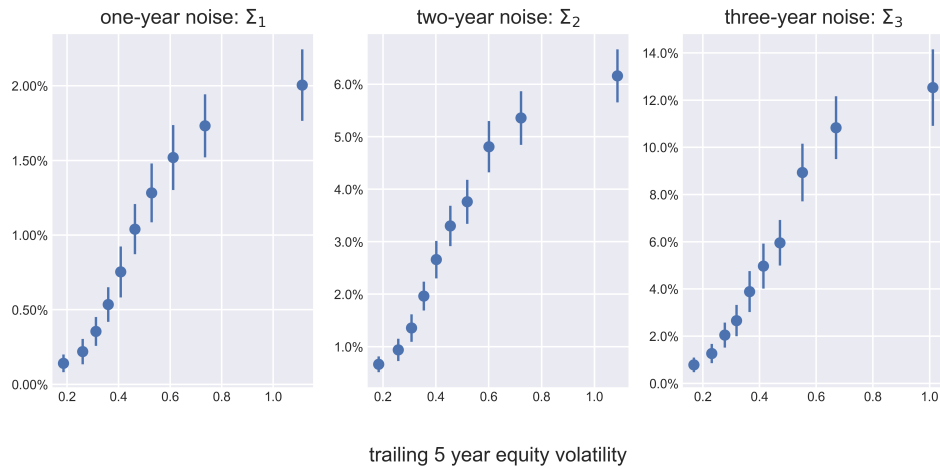
The data features a second relationship: the amount of noise increases with residual earnings volatility. In Figure 5 Panel A we reproduce this fact using return volatility as a measure of earnings surprises. We split the sample into equally-spaced 10 bins of equity volatility (computed using rolling 5-year windows of monthly returns) and compute noise in each one of these bins (see Appendix B for additional details on this calculation). Panel A reports bincatters for each horizon, which illustrate noise is monotonically and smoothly increasing with return volatility.

In Panel B, we check if the model is able to quantitatively match this fact. We do this without assuming an asset pricing model. For each bin of equity returns volatility, we compute earnings conditional volatility as $\sigma_{\epsilon,h}^2 = \text{MSE}_{t+h}^m - \Theta_h$, which is the difference between the MSE of the econometric forecast and the informational advantage. Our model

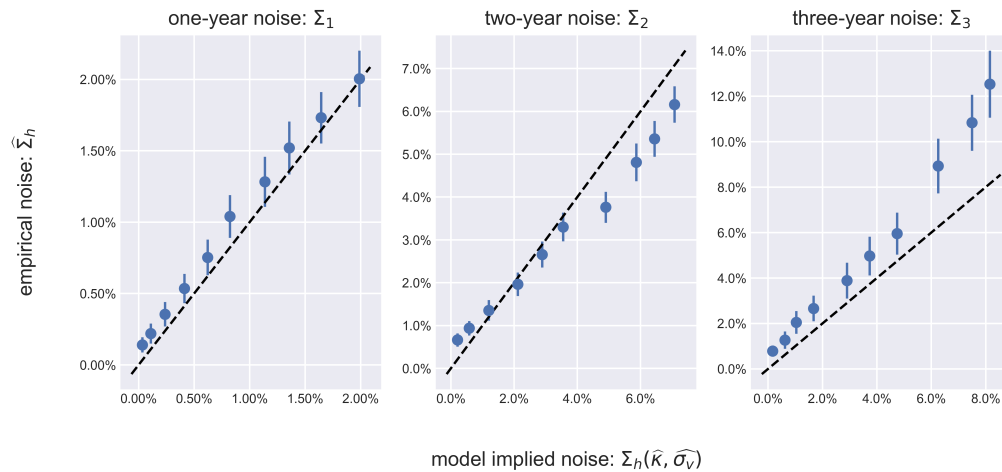
Figure 5. Noise and Volatility

Panel A of this figure plots our estimates of noise at three different horizons across evenly-spaced bins of equity volatility, measured using the annualized standard deviation of monthly stock returns from CRSP over the past five years for firm i in year t . Panel B of this figure plots this same empirical (conditional) noise against the model-implied noise for each bin. We form moments using the Random Forest for statistical forecasts. To estimate conditional noise both Panels, we perform our conditional estimation described in Section 3.5 and Appendix B, replacing N_{it}^h with equity volatility. To estimate the model-implied noise in Panel B, we follow the procedure described in Section 5.2 to generate an estimate of MSE_{t+h}^{FIRE} within each bin, which we convert into an estimate of model implied noise using (12), Proposition 4, and our estimates of κ and σ_v from Table 3. Prior to conducting the estimation, the firm-year level variables in Proposition B1 are trimmed at 5 times the interquartile range, after they are scaled by price. Estimates are normalized by the mean squared EPS across the entire sample. Confidence intervals are calculated using a clustered bootstrap at the firm level with 100 iterations and have 90% joint coverage across all points within each figure. Critical values for these joint confidence intervals are calculated using the 1,000 bootstrap iterations. The final sample used to produce these figures has 47,239, 39,707, and 12,991 firm-year observations at the one-year, two-year, and three-year horizons respectively.

Panel A: Noise versus Equity Volatility



Panel B: Empirical versus Model-Implied Noise



then predicts that:

$$\Sigma_h = \left(\frac{MSE_{t+h}^m - \Theta_h}{\kappa^2 + MSE_{t+h}^m - \Theta_h} \right)^2 \sigma_v^2$$

which is decreasing with κ , the weight of the rational forecast, and increasing with $MSE_{t+h}^m - \Theta_h$, the FIRE mean squared error. In Panel B, we plot empirically estimated noise versus the noise predicted by the above equation, using our estimates of κ and σ_v from Table 3. While our model is statistically rejected at the two and three-year horizons, the fit is quantitatively reasonable. This is impressive since the estimation in Table 3 did not target these cross-sectional moments, but rather the term structure of noise and bias.

6 Conclusion

We study the term structure of subjective expectations using analyst earnings forecasts. We find these forecasts perform better than statistical forecasts at short horizons, but underperform at longer horizons. Decomposing these differences in forecasting accuracy, we find the decreasing accuracy of subjective forecasts with the forecasting horizon is driven by an upward sloping term structure of bias and noise, while the information advantage of subjective forecasters is relatively constant across horizons.

Existing models, in their current form, lack a feature to match these upward sloping term structures of bias and noise. We propose such a mechanism based on bounded rationality and noisy defaults. This model quantitatively matches these term structures and our estimation of it implies a significant degree of noise in forecasters' cognitive defaults. The model is parsimonious, as it has three key parameters: default noise, expectations bias, and the relative weight between rational forecast and noise default. This last parameter succeeds at matching both the term structures of noise and bias.

Our model of belief formation has two direct implications for which we find empirical support: (i) an upward sloping term structure of disagreement and (ii) noise should be an increasing function of earnings volatility. When fitted on the term structures of noise and bias, our model also manages to replicate these two facts quantitatively.

A crucial aspect of our model is the existence of noisy defaults, for which we appeal to existing evidence but give no formal micro-foundation. In this sense, our model provides the reduced-form representation that a more micro-founded model must admit in order to match our empirical results. In subsequent work, we hope to enrich the model in this direction, in

addition to studying more closely the properties of individual-level rather than consensus-level forecasts.

References

- Abadie, Alberto and Maximilian Kasy (2020), “The Risk of Machine Learning.” *Review of Economic Studies*.
- Afrouzi, Hassan, Spencer Kwon, Augustin Landier, Yueran Ma, and David Thesmar (2021), “New Experimental Evidence on Expectation Formation.” *Working Paper*, 1–67.
- Allcott, Hunt and Nathan Wozny (2014), “Gasoline Prices, Fuel Economy, and the Energy Paradox.” *The Review of Economic Studies*, 96, 779–795.
- Andrews, Isaiah, Matthew Gentzkow, and Jesse M. Shapiro (2017), “Measuring the Sensitivity of Parameter Estimates to Estimation Moments.” *The Quarterly Journal of Economics*, 132, 1553–1592q.
- Angeletos, George-Marios, Zhen Huo, and Karthik A. Sastry (2020), “Imperfect Macroeconomic Expectations: Evidence and Theory.” *NBER Macroeconomics Annual*.
- Azaredo da Silveira, Rava, Yeji Sung, and Michael Woodford (2020), “Optimally Imprecise Memory and Biased Forecasts.” *Working Paper*, 85.
- Ball, Ryan T. and Eric Ghysels (2018), “Automated Earnings Forecasts: Beat Analysts or Combine and Conquer?” *Management Science*, 64, 4936–4952.
- Belloni, Alexandre, Victor Chernozhukov, and Christian Hansen (2011), “Inference for high-dimensional sparse econometric models.” *Advances in Economics and Econometrics: Tenth World Congress Volume 3, Econometrics*, 245–295.
- Bianchi, Francesco, Sydney C. Ludvigson, and Sai Ma (2020), “Belief Distortions and Macroeconomic Fluctuations.” *Working Paper*.
- Bordalo, Pedro, Nicola Gennaioli, Rafael La Porta, and Andrei Shleifer (2019), “Diagnostic Expectations and Stock Returns.” *Journal of Finance*, 74, 2839–2874.
- Bordalo, Pedro, Nicola Gennaioli, Yueran Ma, and Andrei Shleifer (2018), “Over-reaction in Macroeconomic Expectations.” *Working Paper*, 1–49.
- Bordalo, Pedro, Nicola Gennaioli, Yueran Ma, and Andrei Shleifer (2020), “Overreaction in Macroeconomic Expectations.” *American Economic Review*, 110, 2748–2782.
- Bouchaud, Jean Philippe, Philipp Krüger, Augustin Landier, and David Thesmar (2019), “Sticky Expectations and the Profitability Anomaly.” *Journal of Finance*, 74, 639–674.
- Bradshaw, Mark T., Michael S. Drake, James N. Myers, and Linda A. Myers (2012), “A re-examination of analysts’ superiority over time-series forecasts of annual earnings.” *Review of Accounting Studies*, 69–76.
- Brown, Lawrence D., Andrew C. Call, Michael B. Clement, and Nathan Y. Sharp (2015), “Inside the ”Black Box” of sell-side financial analysts.” *Journal of Accounting Research*, 53, 1–47.
- Brown, Lawrence D. and Michael S. Rozeff (1978), “The Superiority of Analyst Forecasts as Measures of Expectations: Evidence from Earnings.” *Journal of Finance*, 33, 1–16.
- Bryzgalova, Svetlana, Jiantao Huang, and Christian Julliard (2020), “Bayesian Solutions for the Factor Zoo: We Just Ran Two Quadrillion Models.” *Working Paper*.

- Cao, Kai and Haifeng You (2020), “Fundamental Analysis via Machine Learning.” *Working Paper*.
- Cao, Sean S., Wei Jiang, Junbo L. Wang, and Baozhong Yang (2021), “From Man vs . Machine to Man + Machine: The Art and AI of Stock Analyses.” *Working Paper*.
- Chen, Qi and Wei Jiang (2006), “Analysts’ weighting of private and public information.” *Review of Financial Studies*, 19, 319–355.
- Chernozhukov, Victor, Denis Chetverikov, Mert Demirer, Esther Duflo, Christian Hansen, Whitney K. Newey, and James M. Robins (2016), “Double/Debiased Machine Learning for Treatment and Causal Parameters.” *Working Paper*.
- Chetverikov, Denis, Zhipeng Liao, and Victor Chernozhukov (2020), “On cross-validated Lasso in high dimensions.” *Annals of Statistics*, 40.
- Coibion, Olivier and Yuriy Gorodnichenko (2015), “Information rigidity and the expectations formation process: A simple framework and new facts.” *American Economic Review*, 105, 2644–2678.
- Daniel, Kent, Avanidhar Subrahmanyam, and David Hirshleifer (1998), “Investor Psychology and Security Market Under and Overreactions.” *Journal of Finance*, 53, 1839–1885.
- D’Arienzo, Daniele (2020), “Maturity Increasing Over-reaction and Bond Market Puzzles.” *Working Paper*.
- Dessaint, Olivier, Thierry Foucault, and Laurent Frésard (2020), “Does Big Data Improve Financial Forecasting? The Horizon Effect.” *Working Paper*.
- Enke, Benjamin and Thomas Graeber (2020), “Cognitive Uncertainty.” *Working Paper*.
- Eyster, Erik, Matthew Rabin, and Dimitri Vayanos (2019), “Financial Markets Where Traders Neglect the Informational Content of Prices.” *Journal of Finance*, 74, 371–399.
- Fuster, Andreas, David Laibson, and Brock Mendel (2010), “Natural Expectations and Macroeconomic Fluctuations.” *Journal of Economic Perspectives*, 24, 67–84.
- Gabaix, Xavier (2014), “A Sparsity-Based Model of Bounded Rationality.” *The Quarterly Journal of Economics*, 1661–1710.
- Gabaix, Xavier (2019), “Behavioral inattention.” *Handbook of Behavioral Economics*, 2, 261–343.
- Giannone, Domenico, Michele Lenza, and Giorgio E. Primiceri (2018), “Economic Predictions with Big Data: The Illusion of Sparsity.” *Working Paper*.
- Giglio, Stefano and Bryan T. Kelly (2018), “Excess volatility: Beyond discount rates.” *The Quarterly Journal of Economics*, 133, 71–127.
- Greenwood, Robin and Andrei Shleifer (2014), “Expectations of returns and expected returns.” *Review of Financial Studies*, 27, 714–746.
- Groysberg, Boris, Paul M. Healy, and David A. Maber (2011), “What Drives Sell-Side Analyst Compensation at High-Status Investment Banks?” *Journal of Accounting Research*, 49, 969–1000.
- Gu, Shihao, Bryan T. Kelly, and Dacheng Xiu (2018), “Empirical Asset Pricing via Machine Learning.”

- Hahn, Jinyong (1996), “A note on bootstrapping generalized method of moments estimators.” *Econometric Theory*, 12, 187–197.
- Hansen, Jorge W and Christoffer Thimsen (2020), “Forecasting Corporate Earnings with Machine Learning.” *Working Paper*.
- Harford, Jarrad, Feng Jiang, Rong Wang, and Fei Xie (2019), “Analyst career concerns, effort allocation, and firms’ information environment.” *Review of Financial Studies*, 32, 2179–2224.
- Horowitz, Joel L. (2001), “The Bootstrap.” *Handbook of Econometrics*, 5, 3159–3228.
- Kahneman, Daniel, Olivier Sibony, and Cass Sunstein (2021), *Noise: A Flaw in Human Judgment*.
- Kahneman, Daniel and Amos Tversky (1972), “Subjective probability: A judgment of representativeness.” *Cognitive Psychology*, 3, 430–454.
- Khaw, Mel Win, Ziang Li, and Michael Woodford (2019), “Cognitive Imprecision and Small-Stakes Risk Aversion.” *Working Paper*.
- Kothari, S. P. (2001), “Capital markets research in accounting.” *Journal of Accounting and Economics*, 31, 105–231.
- Kothari, S. P., Eric C. So, and Rodrigo Verdi (2016), “Analysts’ Forecasts and Asset Pricing: A Survey.” *Annual Review of Financial Economics*, 1–23.
- Kozak, Serhiy, Stefan Nagel, and Shrihari Santosh (2020), “Shrinking the cross-section.” *Journal of Financial Economics*, 135, 271–292.
- Kumar, Alok, Ville Rantala, and Ruoxi Xu (2021), “Social Learning and Analyst Behavior.” *Journal of Financial Economics*.
- Manski, Charles F. (2017), “Survey Measurement of Probabilistic Macroeconomic Expectations: Progress and Promise.” *NBER Macroeconomics Annual*.
- Mullainathan, Sendhil and Jann Spiess (2017), “Machine learning: An applied econometric approach.” *Journal of Economic Perspectives*, 31, 87–106.
- Nagel, Stefan (2021), *Machine Learning in Asset Pricing*. Princeton University Press, Princeton and Oxford.
- Patton, Andrew J. and Allan Timmermann (2010), “Why do forecasters disagree? Lessons from the term structure of cross-sectional dispersion.” *Journal of Monetary Economics*, 57, 803–820.
- Patton, Andrew J. and Michela Verardo (2012), “Does beta move with news? Firm-specific information flows and learning about profitability.” *Review of Financial Studies*, 25, 2789–2839.
- Satopää, Ville, Marat Salikhov, Philip E. Tetlock, and Barb Mellers (2020), “Bias, Information, Noise: The BIN Model of Forecasting.” *Working Paper*.
- Schmidt-Hieber, Johannes (2020), “Nonparametric regression using deep neural networks with relu activation function.” *Annals of Statistics*, 48, 1875–1897.
- Shumway, Tyler (1997), “The Delisting Bias in CRSP Data.” *Journal of Finance*, 52, 327–340.

- Sims, Christopher A. (2003), “Implications of rational inattention.” *Journal of Monetary Economics*, 50, 665–690.
- So, Eric C. (2013), “A new approach to predicting analyst forecast errors: Do investors overweight analyst forecasts?” *Journal of Financial Economics*, 108, 615–640.
- van Binsbergen, Jules H., Xiao Han, and Alejandro Lopez-Lira (2020), “Man vs. Machine Learning: The Term Structure of Earnings Expectations and Conditional Biases.” *Working Paper*.
- van Wieringen, Wessel N. (2020), “Lecture notes on ridge regression.” *Working Paper*.
- Wager, Stefan and Susan Athey (2018), “Estimation and Inference of Heterogeneous Treatment Effects using Random Forests.” *Journal of the American Statistical Association*, 113, 1228–1242.
- Woodford, Michael (2003), “Imperfect Common Knowledge and Monetary Policy.” In *Knowledge, Information, and Expectations in Modern Macroeconomics: In Honor of Edmund S. Phelps*, 25–58, Princeton University Press, Princeton, NJ.
- Woodford, Michael (2020), “Modeling imprecision in perception, valuation, and choice.” *Annual Review of Economics*, 12, 579–601.